"The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723205"

Deliverable D2.4

Evaluation of new human-centred low-cost measures

Due date of deliverable: 31/12/2019
Actual submission date: 31/12/2019

Authors:

Annika Dressler, DLR
Anne Silla, VTT
Grigore M. Havârneanu, UIC
Neofytos Boufidis, CERTH-HIT
Virginie Taillandier, SNCF
Aida Herranz, FFE
Jan Grippenkoven, DLR

Project details

Project acronym	SAFER-LC
Project full title	SAFER Level Crossing by integrating and optimizing road-rail infrastructure management and design
Grant Agreement no.	723205
Call ID and Topic	H2020-MG-2016-2017, Topic MG-3.4-2016
Project Timeframe	01/05/2017–30/04/2020
Duration	36 Months
Coordinator	UIC – Marie-Hélène Bonneau (bonneau@uic.org)

Document details

Title	Evaluation of new human-centred low-cost measures
Workpackage	WP2
Date of the document	28/12/2019
Version of the document	10
Responsible partner	DLR
Reviewing partner	VTT, UIC, CERTH-HIT, INTADER, SNCF, FFE
Status of the document	Final
Dissemination level	Public

Document history:

Revision	Date	Description
01	04/09/2019	First Draft
02	27/09/2019	Version 2, including changes discussed in PM8
03	06/12/2019	Version 3 for review, including all results tables
04 – 08	06 – 16/12/ 2019	Intermediate working versions
09	16/12/2019	Version 9 for final review by partners
10	29/12/2019	Final version
11	31/12/2019	Ethical and final review

Consortium - List of partners

Partner No	Short name	Name	Country
1	UIC	International Union of Railways	France
2	VTT	Teknologian tutkimuskeskus VTT Oy	Finland
3	NTNU	Norwegian University of Science and Technology	Norway
4	IFSTTAR	French institute of science and technology for transport, development and networks	France
5	FFE	Fundación Ferrocarriles Españoles	Spain
6	CERTH-HIT	Centre for Research and Technology Hellas - Hellenic Institute of Transport	Greece
7	TRAINOSE	Trainose Transport – Passenger and Freight Transportation Services SA	Greece
8	INTADER	Intermodal Transportation and Logistics Research Association	Turkey
9	CEREMA	Centre for Studies and Expertise on Risks, Environment, Mobility, and Urban and Country planning	France
10	GLS	NeoGLS	France
11	RWTH	Rheinisch-Westfaelische Technische Hochschule Aachen University	Germany
12	UNIROMA3	University of Roma Tre	Italy
13	COMM	Commsignia Ltd	Hungary
14	IRU	International Road Transport Union - Projects ASBL	Belgium
15	SNCF	SNCF	France
16	DLR	German Aerospace Center	Germany
17	UTBM	University of Technology of Belfort-Montbéliard	France

Executive summary

This deliverable describes the methods applied and the results achieved during the second phase of Task 2.3 in the SAFER-LC project: the evaluation of new human-centred low-cost measures to improve safety at level crossings (LCs). The European project SAFER-LC – Safer level crossing by integrating and optimizing road-rail infrastructure management and design – aimed to improve safety in road and rail transport by minimising the risk of LC accidents, focusing on both technical solutions and human processes. Within the project, the objective of Work Package 2 (WP2) was to enhance the safety performance of level crossing infrastructures from a human-factors perspective, making them more self-explaining and forgiving.

Task 2.3 specifically aimed to design human-centred low-cost countermeasures to enhance the safety of current LC infrastructures and, in a later step, to evaluate these countermeasure designs from a human-factors perspective. This objective was driven by the insights of the major role that road user behavior plays in accidents at level crossings and the need for safety measures to be affordable to enable their application to a large number of crossings and the achievement of tangible safety effects. The activities in the design of countermeasures were performed in the first phase of task 2.3 from May 2017 to October 2018. They resulted in a list of 89 reviewed LC safety measures, of which 36 measures were for use at passive LCs, 29 for LCs with barriers, and 24 for use at all kinds of LCs. For the purpose of evaluation, Task 2.3 referred to two main inputs from other tasks within SAFER-LC: the human factors methodological framework developed in Task 2.2 and the pilot tests of innovative LC safety measures performed in Work Package 4 (WP4).

The human factors methodological framework was developed to define what aspects of human behavior should be considered when trying to assess the suitability of a LC safety measure. It also defined important context variables that influence this suitability, including environmental factors such as LC type, layout, weather, traffic etc. as well as the issue of acceptance by different stakeholders. The methodological framework is based on sociotechnical systems theory, relevant models of human cognition and behavior, and analytical tools and empirical approaches from related research projects. Its development resulted in the definition of three sets of criteria important to the human factors assessment of a given LC safety measure. To facilitate and structure the application of the framework, a human factors assessment tool (HFAT) was developed. Its core is a survey comprising checklists and forms to assess the three sets of criteria defined. The tool helps to collect and systemize relevant information on a given LC safety measure in order to enable a reasoned estimation of its effects in road user behavior, user experience and social perception.

The pilot tests in WP4 involved two kinds of tests. One kind focused on demonstrating the feasibility of technical solutions to improve LC safety. The other one was concerned with the effects of LC safety measures on road user behavior. This included two simulator studies of infrastructural safety measures, an online survey based on videos of a train-mounted countermeasure in a real rail environment, a field test of an in-vehicle LC proximity warning, and a field test of two infrastructural measures. Based on the results of these tests, the pilot site leaders used the HFAT to assess the piloted measures from a human-factors perspective. Using the HFAT enabled the presentation of the results in a common format, although the input studies used different methods and measured different indicators.

The LC safety measures evaluated in this way were: *blinking lights for the locomotive front*, *coloured road markings on approach to the LC*, *in-vehicle proximity warning*, *rings upstream of the LC*, *traffic light*, *blinking amber light with train symbol*, *funnel effect pylons*, *message , "← Is a train coming? →" written on road*, *peripheral blinking lights*, *rumble strips*, *sign "← Is a train coming? →"*, and *speed bump and flashing posts.*

The four measures assessed to most facilitate safe road user behavior in the HFAT evaluation were the *blinking lights for the locomotive front*, the two *in-vehicle proximity warnings,* and the *peripheral blinking lights*. Minding the evidence collected in the HFAT, this assessment is rather certain for the two measures involving blinking lights, and more tentative for the *in-vehicle proximity warnings.* Stakeholder acceptance and user trust are expected to be sufficient to allow for successful implementation of these measures, minding the principles of stakeholder participation and user-friendly design.

Two measures scored particularly low on the assessment of behavioral safety effects: the *funnel effect pylons* and the *message "← Is a train coming? →" on the road*. Both assessments are tentative, as the findings from the pilot are the only evidence available by now. Due to the low expected efficacy, acceptance and trust values were not considered in these cases.

The seven remaining measures were attested a medium effectivity on the facilitation of safe behavior. These assessments are more certain for the *rumble strips* and the *sign "← Is a train coming? →",* and remain tentative for the *coloured road markings on approach to LC*, *the rings upstream of the LC*, the *traffic light*, the *blinking amber light with a train symbol*, and the *speed bumps and flashing posts* due to the limited availability of evidence. Based on the acceptance and trust values obtained with the HFAT, successful implementation appears possible for most of these measures. Some difficulty in implementation is expected based on the acceptance assessment for the *coloured road markings on approach to LC*, the *rings upstream of the LC*, and the *funnel effect pylons*.

Beyond its use as a tool to guide and evaluate empirical research on LC safety, the HFAT can also be used by road and railway transport stakeholders as a checklist to support the consideration of human factors aspects in the evaluation of LC safety measures. Using the HFAT in this function can help to assess the suitability of a LC safety measure to different railway environments and user requirements and to avoid efficacy barriers, by considering the important issues of acceptance and social perception of road users and other stakeholders.

The results obtained in SAFER-LC Task 2.3, the design and evaluation of human-centered low-cost measures to improve LC safety, will be used as one main input in the implementation of the *SAFER-LC toolbox*, a web-based tool for anyone concerned with LC safety. The toolbox is conceived to be a guide to best practice that integrates all the recommendations, promising interventions, and specifications developed during the project with the empirical evidence collected from the scientific literature and the pilot tests. The toolbox will be accessible free of charge at the end of the project and will continue to be maintained, updated and improved by the International Union of Railways (UIC) for the benefit of the road- and railway-safety community.

Table of content

List of abbreviations

D	Deliverable (related to project SAFER-LC)
GNSS	Global navigation satellite system
HF	Human factors
HFAT	Human factors assessment tool
LC	Level crossing
LED	Light-emitting diode
MRU	Motorized road users
ROI	Region of interest
T	Task (related to project SAFER-LC)
VRU	Vulnerable road users
WP	Work package (related to project SAFER-LC)

1. INTRODUCTION

1.1. Purpose and structure of the document

This deliverable describes the methods and results of the second phase of Task 2.3 in the SAFER-LC project: the evaluation of new human-centred low-cost measures to improve safety at level crossings (LCs). For this purpose, a short introduction to the scope and objectives of the project is given in section 1, followed by a presentation of the specific objectives of Task 2.3. Section 1 closes with a short introduction to the LC safety measures that were evaluated and the kind of tests they were subjected to in WP 4 of the SAFER-LC project. Section 2 introduces the methods used in the human factors evaluation of measures, especially the *Human Factors Assessment Tool* that was developed for this purpose, and provides a more detailed overview of the measures evaluated. The results of the evaluation are presented in section 3, containing a comparison of the measures on their applicability to different types of LCs and circumstances, the assessment of the measures' suitability to support road users in behaving safely at LCs at different stages of human information processing, and estimations of the measures' acceptability and perceived reliability on the part of road users and other stakeholders. Sections 4 and 5 contain reflections on the process and an outlook on the further exploitation of the results in the form of an open-access toolbox.

1.2. Background and objectives

1.2.1. The SAFER-LC project

The European project SAFER-LC (Safer level crossing by integrating and optimizing road-rail infrastructure management and design) aims to improve safety in road and rail transport by minimising the risk of LC accidents. This is done by developing a cross-modal set of innovative solutions and tools for the proactive management of LC safety and by developing alternatives for the future design of level-crossing infrastructure.

The solutions and tools that are in development in the SAFER-LC project seek to enable road and rail stakeholders to find more effective ways to: (1) detect potentially dangerous situations leading to collisions at level crossings, (2) prevent incidents by innovative user-centred LC design, and (3) mitigate the consequences of disruptions due to accidents or other critical events. The main output of the SAFER-LC project is a web-based toolbox accessible through a user-friendly interface that integrates the project's practical results, tools and recommendations to help both rail and road stakeholders to improve safety at LCs.

The project focuses both on technical solutions, such as smart detection services and advanced infrastructure-to-vehicle communication systems and on human processes to adapt infrastructure designs to road user needs and to enhance coordination and cooperation between different stakeholders from different land transportation modes. A further challenge is to demonstrate the acceptance of the proposed solutions by both rail and road users and to implement the solutions cost-efficiently.

Within the project, the objective of Work Package 2 (WP2) is to enhance the safety performance of level crossing infrastructures from a human factors (HF) perspective, making them more self-explaining and forgiving.

1.2.2. **Design and evaluation of human-centered low-cost measures**

Objectives

The two main objectives of SAFER-LC Task 2.3 were to (1) design concepts of human-centred low-cost countermeasures to enhance the safety of current LC infrastructures and (2) to evaluate these countermeasure designs from a human-factors perspective.

The first objective was to be achieved by identifying knowledge gaps, new approaches and out of the box ideas concerning LC design and, on this basis, proposing new technological and non-technological measures to enhance LC safety. The process was to be inspired by the insights gathered on the state of the art of LC design and measures in previous work in SAFER-LC (WP1 and WP2), the consultation of experts from road and rail transport, and the findings of worldwide research in the field of human factors applied to level crossing safety. The conception and selection of promising countermeasures were envisaged to encompass entirely new ideas as well as upgrades of existing measures to enhance their innovation potential as well as their self-explaining and forgiving nature. The activities in the design of countermeasures were performed in the first phase of task 2.3 from May 2017 to October 2018. The process and results are documented in detail in Deliverable 2.3 (Dreßler, Silla, Kortsari, Havârneanu, Whalley, Lorenzo & Grippenkoven, 2018). A short summary is given in section 1.3.

The second objective encompassed the evaluation of proposed measures to understand their effectiveness in enhancing LC safety, using the criteria defined in the human factors methodological framework developed in Task 2.2 and involving the results from the empirical tests at the project test sites performed as part of WP4 (see SAFER-LC Consortium, 2017). The activities in the evaluation phase and their results are documented in the present deliverable.

Definitions and specification of measures to be addressed

Summing up the specifications for countermeasures to be designed and collected according to the objectives of Task 2.3, the proposed measures to enhance LC safety were to be (1) human-centred, (2) low-cost and (3) new or innovative, and they were supposed to work by making LC infrastructures more (4) self-explaining and (5) forgiving. For a common understanding, these five concepts were further defined as follows:

(1) *Human-centred* measures are measures whose effect is achieved by influencing road user behaviour at LCs, especially by enhancing adaptive behaviour (e.g. looking for a train before crossing instead of crossing without looking; waiting in front of barriers instead of trying to circumvent them). While the term could also refer to other human agents in the railway system (train drivers, signallers, workers on the tracks, etc.), these were excluded from the scope of the task. The focus on measures that influence road user behaviour was chosen as the vast majority of LC accidents is caused by maladaptive behaviour on the side of road users (DB Netze, 2016; Grippenkoven & Dietsch, 2015; Grippenkoven, 2017).

(2) *Low-cost* measures are measures that cost less than a classic upgrade (e.g. equipping a formerly passive LC with half-barriers; installing full-barriers at a former LC with half-barriers) when applied to a large number of LCs.
(3) *New* or *innovative* measures are measures that are not already in common use to protect LCs in the European countries.
(4) *Self-explaining* refers to the clear and appropriate design of safety measures implemented at the LC which supports adequate situation awareness, meaning that it supports (1) the detection and perception of the situation; (2) the understanding the meaning of signs and measures; and (3) the ability to project the current status of the traffic situation at the level crossing into the future (Havârneanu, Silla, Whalley, Kortsari, Dreßler, & Grippenkoven, 2018).
(5) *Forgiving* means that the safety measures implemented at a LC include appropriate measures to counteract road user misbehaviour (e.g. errors, violations, or deficient behavioural adaptation), and if misbehaviour occurs, the system is able to mitigate the consequences (Havârneanu et al., 2018).

Furthermore, taking into account the infrastructure focus expressed in the objectives, the task was specified to focus on measures that can be applied or have a direct effect on road user behaviour at the level crossing itself. This includes, for instance, traffic infrastructure elements that can be installed at a crossing (e.g. road elements like speed bumps or lane dividers, light markings, signs), law enforcement measures noticeable at the LC (e.g. cameras) as well as changes in operational procedures that lead to a direct change of the situation at the crossing (e.g. shortening and equalizing closure times by adapting the timing of closure to the speed of the respective train). Measures that are not applied at the crossing itself (e.g. the revision of driving education) were excluded from the focus of the task.

<u>Main issues concerning road user behaviour at different LC types</u>

Taking a human-factors perspective on the action of encountering and crossing a level crossing, there are five steps of information processing that road users need to complete for the purpose of a safe traverse (Graab, Donner, Chiellino, & Hoppe, 2008; Grippenkoven, 2017; Havârneanu et al., 2018):

(1) to *detect* at least parts of the safety layout of a level crossing (e.g. signs),
(2) to correctly *identify* the kind of level crossing that these parts of the safety layout belong to,
(3) to *retrieve* schemas and scripts connected to passing the LC from memory (or other sources),
(4) to *decide* on an appropriate action, i.e. to form an intention that matches the current situation, and finally
(5) to properly *execute* the intended action.

Things can go wrong in each of these stages, leading to errors or violations. A wide range of environmental factors as well as individual conditions and traits influence the probability of errors and violations occurring at LCs (e.g. sight distances, road layout, weather conditions, road and rail traffic density; reduced sensory or motor capacities, distraction, time pressure, fatigue or individual propensities in risk assessment). However, when considering the *quality* of maladaptive road user behaviour that LC safety measures seek to tackle, there is one factor exerting a major influence: the type of protection applied at the respective crossing. The reason for the crucial role of the protection type is that it is the primary determinant of behavioural demands imposed upon road users in the aforementioned phases of information processing after LC detection. The most basic distinction concerning these behavioural demands depends on the presence or absence of active controls and barriers at the LC. Closed barriers represent a strong and almost impossible-to-misunderstand cue to road users that they should stop in front of the crossing. In contrast, on approach to passive LCs, road users need to determine on their own whether they need to stop and grant the right of way to an approaching train, and therefore must enter into another loop of visual search and potential detection after detecting the crossing itself. Thus, there are fundamental differences in the action schemata that need to be activated and executed facing passive LCs compared to LCs that are equipped with barriers (Grippenkoven & Dreßler, 2018).

The differences in behavioural demands are associated with differences in the main issues that arise in road user behaviour at passive and active LCs, respectively, and that need to be defined as the target of safety measures. The focus of problematic behaviour at passive LCs is an insufficient preparation of the traversing action in terms of obtaining information and putting oneself in a position to stop in good time if necessary. While these aspects are much easier to accomplish with the help of active signals and barriers at active LCs, the main challenge for road users at this type of crossing is the extrinsic imposition of waiting time that comes into conflict with the individual's mobility goals and potentially provokes violations (Seehafer, 1997). Apart from these motivationally induced issues, problems with anticipatory action planning could lead to a situation in which a road user gets stuck on the rails or "trapped" within the barriers because of a wrong estimation or omitted consideration of the time needed to cross (e.g. due to traffic tailback or an overestimation of their own achievable speed in light of an imminent closure; Pelz, 2011). Though this problem is accentuated by the presence of barriers that represent an additional obstacle in leaving the tracks, it equally needs to be considered at passive LCs.

Table 1 summarizes the challenges in road user behaviour observed at passive and active LCs that were used as a basis for the design thinking and search for safety measures in Task 2.3.

Apart from the differences that have been pointed out, passive and active LCs also have a number of things in common in terms of how road users can be supported in successfully dealing with them. Looking at the first stage of information processing, LC detection, measures that enhance the conspicuity of the LC will be beneficial in either case. In the following stages, although the specific design needs to be adapted to the demands at the respective crossing type, common measures could be used to support them. For example, the activation of the correct action scheme (either to slow down and look left and right for a train or to watch the status of active controls and take heed of the signals) could be helped by providing cues and information through the same channel (e.g. an in-vehicle information system). Furthermore, similar measures could be applicable at either kind of LC to support road users in refraining from entering the track area when they cannot be sure they can leave it in good time (Cale, Gellert, Katz, & Sommer, 2013).

The specific demands and challenges that road users face when approaching and crossing the different kinds of LCs were the starting point of the development and organization of human-centered low-cost safety measures in Task 2.3. In the development process, three basic LC types were distiguished: active LCs with full barriers, active LCs with half-barriers (including other types of active protection that can be circumvented with relative ease) and passive LCs.

Table 1. Challenges with road user behaviour at passive vs. active level crossings

Passive LCs	▪ Insufficient visual scanning of tracks for train ▪ Insufficient adaption of approach speed to scanning needs and potential requirement to stop ▪ Getting stuck on the rails
Active LCs (full-barriers, half-barriers, lights)	▪ Circumventing closed barriers (climbing over / below; swerving around half-barriers) ▪ Passing the LC in spite of active light signals (e.g. flashing red light) ▪ Passing the LC after pre-signalling has begun / while barriers are closing ▪ Getting stuck on the rails

For the purpose of documentation and based on the results obtained, the grouping of measures was rearranged. All measures that could be applied to all LC types were grouped together. The same was done for full- and half-barrier LCs, as just under half of the measures identified for either of these types could also be applied to the other one. Therefore, the results were structured under three basic application categories: (1) passive LCs, (2) active LCs with barriers and (3) all LCs.

1.3. Design of human-centered low-cost measures

The objective of designing human-centered low-cost measures to enhance LC safety was driven by two important insights: (1) the major role that road user behavior plays in accidents at level crossings and (2) the need for safety measures to be affordable to enable their application to a large number of crossings and thus the achievement of tangible safety effects (DB Netze, 2018; European Union Agency for Railways, 2019; Grippenkoven, 2017; Grippenkoven & Dietsch, 2015; UNECE, 2016).

The identification of human-centered low-cost LC safety measures comprised two main phases (see Figure 1): First, a large pool of design ideas was collected from multiple sources. Second, the measures collected were reviewed and filtered based on multiple criteria to come up with a selection of practicable measures with high potential to improve LC safety.

The three sources used in the collection and invention of measures were: a *comprehensive review of international studies, reports and experiences* in the field of human factors applied to LC safety; the *study and discussion of human factors and psychological models* to support the understanding and prediction of road user behavior at LCs; and a *design workshop with road and rail experts* to conceive innovative measures that make level crossings safer by positively influencing road user behavior. The collection and invention activities resulted in a pool of potential countermeasures to enhance LC safety with 185 entries. The pool contained a broad variety of measures in terms of scope, innovation, effect mechanism, technology use, target group, degree of elaboration, status of development and availability of empirical evidence on their effectivity. Moreover, the similarity between a number of the ideas called for a redundancy check. Therefore, the aim of the subsequent selection phase was to review, sort and rank the measures, to identify those recommended for further empirical testing.

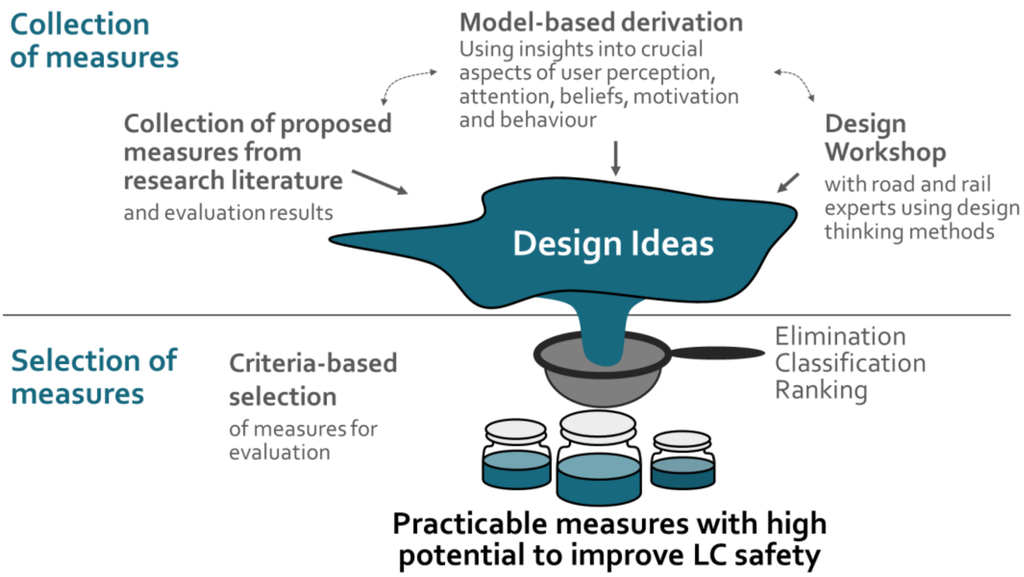


Figure 1. Approach in the definition of new human-centered low-cost countermeasures (Dreßler et al., 2018, p. 15)

The criteria-based selection of measures contained three steps. First, the collection was reviewed in order to eliminate redundancy and exclude measures that did not match the defined scope. The preparation of this step included, among other things, the collection of information from all the projected WP4 test sites on the testability of measures at a given site with respect to effects on road user behaviour. Second, all remaining measures were classified in terms of their application context and effect mechanisms. Third, the measures were assessed on their prospects for reducing accident risk and their need to be further researched. The list of measures was sorted in descending order by, first, the assessed prospects to reduce accident risk and, second, the assessed need for further research, and a rank variable was created based on the resulting order. The result was a list of 89 LC safety measures in which the measure rated with the highest prospects for accident risk reduction ranked first, while measures with lower prospects for risk reduction appeared in a later position. If two measures scored equally on this criterion, the one with the higher need for research ranked higher than the other. Overall, 36 measures were identified for use at passive LCs, 29 for active LCs with barriers, and 24 for use at all kinds of LCs.

1.4. Overview of the measures and tests

The survey of capacities for testing LC safety measures with a focus on road user behavior yielded five test sites where this could be done for at least one or more of the measures in the list:

- a driving simulator environment at Chalon-sur-Saône (France, SNCF; cf. Annex A1),
- a driving simulator environment at Braunschweig (Germany, DLR; cf. Annex A2),
- a real railway environment at Sääksjärvi (railway section closed for piloting), the tests at which were complemented with a user questionnaire (Finland, VTT; cf. Annex A3),

- an urban real-traffic environment with 29 LCs at Thessaloniki (Greece, CERTH-HIT, TRAINOSE; cf. Annex A4),
- a real-world LC at Braunschweig (Germany, DLR; cf. Annex A5).

An additional test site was planned in Turkey, but as the project developed, it had to be put on hold in April 2019 due to political reasons external to the consortium.

It was obvious that not all of the 89 measures could be empirically tested in the course of the project. Some of the measures to be piloted were already defined by the project description. To define further tests, the concerned test site leaders proposed how many and what measures from the list they could test at their sites based on feasibility, capacity and individual appraisal. Finally, 13 measures were subjected to behavioral testing and human factors assessment. Table 2 contains an overview of these measures and the methods used in their primary assessment within the WP4 pilot tests. For each measure, the table also lists the main LC type that the measure was conceived for in the design process, which could either be passive LCs (*passive)*, active LCs with barriers (*active)* or all kinds of LCs (*all;* cf. section 1.3*).* The measures in the table are sorted for this application context, and in alphabetical order within a given context.

As the methods of the different pilot tests have been described in detail in D4.4 (Silla, Lehtonen et al., 2019), these descriptions are not to be repeated in the running text of this deliverable. They can however be found in the annexes A1 to A5 to this document. The overview in Table 2 shows that, overall, 10 of the measures were tested in a simulator environment, three in a field environment, and one both in a simulated and a field environment. The entries in the last column indicate the Annexes in which the methods of the respective pilot are described.

In order to create a common human factors metric that allows to jointly assess the results achieved in the different pilot tests, these results were used as inputs to the human factors assessment method that is described in the following sections.

Table 2. Overview of the human-centred low-cost measures evaluated and assessment methods used in the pilot tests

Measures tested	Main target LC type	Test Methods applied		Methods description in annex
		Simulator Test	Field Test	
Blinking Lights for Locomotive front	All	X	X	A2, A3
Coloured road markings on approach to LC	All	X		A1
In-vehicle proximity warning (1)	All		X	A4
In-vehicle proximity warning (2)	Active	X		A1
Rings upstream of the LC	Active	X		A1
Traffic light	Active	X		A1
Blinking amber light with train symbol	Passive		X	A5
Funnel effect pylons	Passive	X		A1
Message "← Is a train coming? →" written on road	Passive		X	A5
Peripheral blinking lights	Passive	X		A2
Rumble strips	Passive	X		A2
Sign "← Is a train coming? →"	Passive	X		A2
Speed bump and flashing posts	Passive	X		A1

2. METHODS

2.1. Human Factors Evaluation by the Human Factors Assessment Tool

When trying to identify LC safety measures with a high potential to positively influence road user behavior in a given LC context, practitioners as well as researchers are confronted with the problem of assessing and comparing the suitability of different measures in a situation where empirical evidence is only partly available. Moreover, the evidence available often focuses on isolated aspects instead of providing a comprehensive view and can hardly be compared across measures tested in different studies because the methods and indicators used are not equivalent.

The aim of task 2.2 within the SAFER-LC project was to develop a human factors methodological framework that helps to define what aspects of human behavior should be considered when trying to assess the suitability of a LC safety measure. Moreover, the framework should point out what important context variables influence this suitability, including environmental factors such as LC type, layout, weather, traffic etc. as well as the issue of acceptance by different stakeholders. This methodological framework was developed on the basis of sociotechnical systems theory, relevant models of human cognition and behavior, and analytical tools and empirical approaches from related research projects. The development resulted in the definition of three sets of criteria important to the human factors assessment of a given LC safety measure (see Figure 2; Havârneanu, Silla, Whalley, Kortsari, Dreßler, & Grippenkoven, 2018).

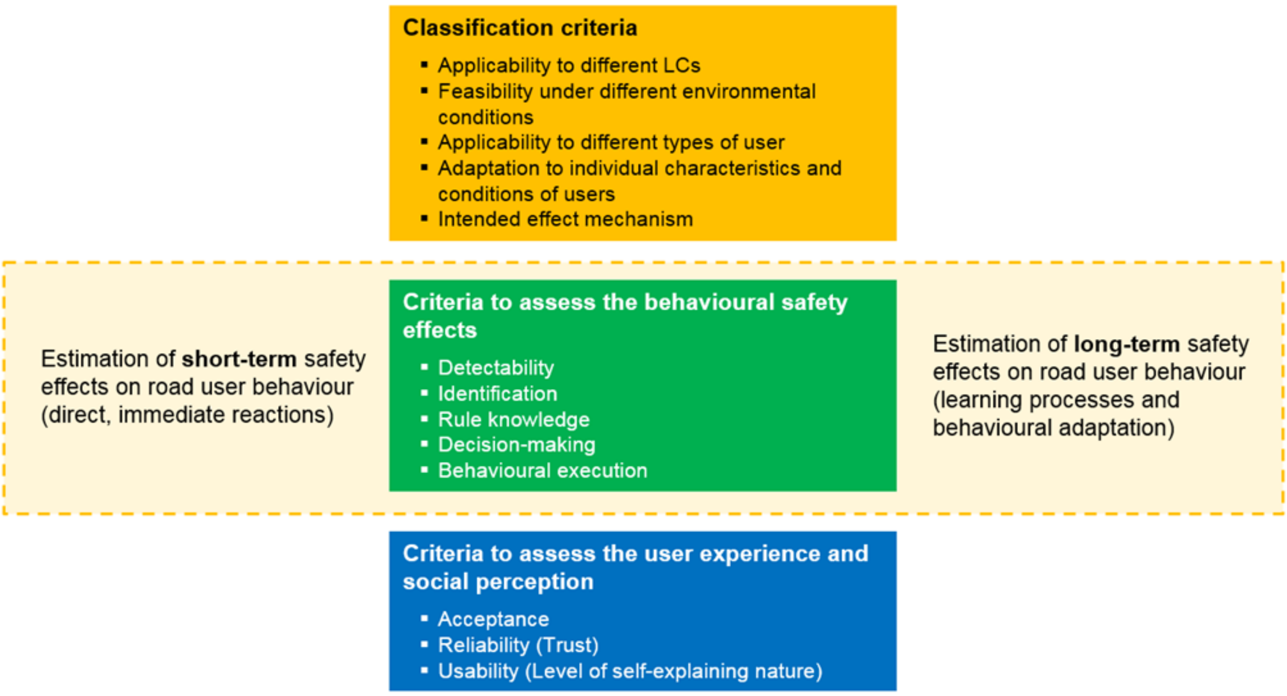


Figure 2. Overview of the sets criteria for HF assessment of LC safety measures.

The first set (orange) consists of classification criteria that describe the scope of the measure under assessment. These criteria specify the compatibility of the measure with different LC and environmental conditions as well as its applicability to different LC user types and characteristics. Moreover, they classifiy the intended effect mechanism via which the measure is expected to affect the road and railway safety. These qualitative criteria are used to define the context and environment in which the safety measure is expected to be effective.

The second set of criteria (green) is used to assess the effects of safety measures on road user behaviour. These criteria are categorized according to the area of psychological function involved. They help estimate changes in road user behaviour (both short and long-term) based on KPIs, literature, expert assessment, LC statistics etc.

The third set of criteria (blue) aims to assess user experience and social perception in terms of the safety measure. The indicators refer to the subjective opinions, and thus this information is collected through a questionnaire among relevant stakeholders and road users or through interviews with selected representatives of these groups.

First and foremost, the methodological framework was developed for use within the SAFER-LC project with the purpose of guiding the empirical testing of human-centered LC safety measures – i.e. the setup of the pilot tests in WP4 – and allowing a detailed evaluation of a given measure from a human-factors point of view. In addition, the methodological framework facilitates the integration of results from different studies done in the project as well as existing research results from the literature.

To facilitate and structure the application of the human factors methodological framework, a human factors assessment tool (HFAT) was developed. Its core is a survey comprising checklists and forms to both collect and systemize relevant information on the LC safety measure under review in order to enable a subsequent reasoned estimation of its effects in road user behavior, user experience and social perception. In the HFAT section concerned with the collection of empirical evidence on safety effects in road user behavior (green), an additional distinction is made between short-term and long-term effects in order to inform the quantitative estimation of the behavioral safety effect. More detailed information on each of the sets of criteria is given in connection with the presentation of the results in section 3. The complete forms of the HFAT are included in Annex A6.

2.2. Measure and test profiles

In the following sections, each of the piloted safety measures is presented in a one-page profile. Each profile is a table aiming to introduce, in a short and common format, the measure's properties and the circumstances under which it was tested in the pilots. The profiles consist of two parts as regards contents. The first part is concerned with the general concept of the measure. Here, each measure is introduced with a general description (*Measure description*) and visualized with at least one illustration. The illustrations mostly show the pilot implementation of the measure, but some are also conceptual or show other implementations. All illustrations are to serve as examples of how the measure could look like. Next, the general measure scope is defined in terms of what road users it aims to address (*Target road users*: *All*, *VRU, or MRU*), what LC types it is mainly conceived for (*Target LC type*: *All*, *Passive*, or *Active*) and how it is assumed to attain its effects on road user behavior under two different perspectives (*Proposed effect mechanism*, *Main psychological function involved*). The proposed effect mechanism categories are based on a taxonomy used by Silla, Seise and Kallberg (2015) which was further adapted for the SAFER-LC project. They contain the following mechanisms: *improves train detection, improves LC detection, controls access to or supports egress from LC, reduces the approach speed of vehicles, increases awareness of correct behavior and consequences of violation, makes waiting time more tolerable, improves physical environment of LC, provides up-to-date information about LC status, supports LC safety actions, improves the possibilities of VRU to cross the LC safely*, and *other*. The classification of the main psychological function involved in the measure's effect followed the theoretical model of human information processing at LCs (Grippenkoven and Dietsch, 2015; Grippenkoven, 2017; Havârneanu et al., 2018), with the stages of *detection* (focus on visual / auditory perception), *identification* (focus on attention and workload), *rule knowledge* (focus on the activation of relevant knowledge), *decision-making* (focus on risk-perception, subjective judgment, and motivational factors), and *execution* (focus on motor execution of action). All taxonomies used to define the scope of a measure are based on the work in the definition of new human-centered low-cost countermeasures (cf. Dreßler et al., 2018).

The second part of the profiles is concerned with the specific form in which each of the measures was implemented and tested in the pilot. This includes the technical description of the most relevant features (e.g. distance from the LC at which the measure was applied, parameters used for blinking lights, etc.), to enable a better understanding and assessment of the exemplary implementation (*Measure specification in test*). As a counterpart to the general measure scope, the *test scope* specifies the application context that was addressed by the test method (*Road user role in test*, *LC type in test*). It includes information on the test environment (*Type of test:* Simulation, Field test, Online survey based on field-test videos) and gives a reference to the annex in which a detailed methods description of the respective test can be found (A1 to A5). Finally, the kind of data that was measured (*Data assessed*) is given as a counterpart to the theoretical effect mechanisms and psychological functions involved. The order in which the measures are presented follows the structure introduced in section 1.4, starting with the measures for use at all LC types and continuing with the measures for use at active LCs with barriers and passive LCs.

2.2.1. Measures for all LC types

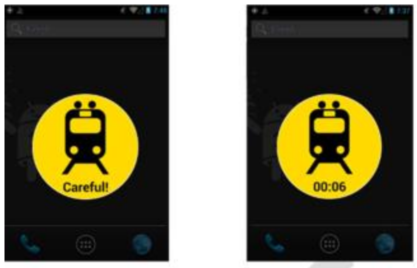
Blinking Lights for Locomotive front

Illustration		
Measure description	Improvement of train detectability using blinking lights. Additional strobe lights on the locomotive are to exogenously capture road users' visual attention, as road users often do not look out for a train ahead of a LC (Grippenkoven & Dietsch, 2015). The automatic shift of attention elicited by flickering stimuli in the periphery of the visual field is an autonomous physiological reaction (Yantis, 2000). It causes road users to visually orient towards the stimuli without requiring a self-initiated visual search.	
Measure Scope	**Target road users**	**Target LC type**
	All road users.	All LCs (most effective at passive)
	Proposed effect mechanism	**Main psychological function involved**
	Improves train detection	Detection
Measure specification in test 1	Three blinking lights were implemented on the locomotive according to the prevailing regulations for train lighting, below the three headlights. To enhance detectability also from the side, the two bottom lights stretched around the corner from the front to the side, and there was an additional blinking top element at the side (see illustration). The blinking was triggered when the train was at around 300 m ahead of the LC and stopped when the train had passed the LC. It occurred in continuous cycles with a frequency of 6.25 Hz (0.16 s off/on).	
Test 1 Scope	**Road user role in test**	**LC type in test**
	Car driver	Passive
	Type of test	**Data assessed**
	Driving simulation (A2)	Gaze, driving dynamics, subjective assessments
Measure specification in test 2	The additional warning lights were installed to the train according to the prevailing regulations (e.g. below the head lights). The test equipment contained three high intensity LED lights and control unit. LED lights were high beam accessories and accepted to be used in road traffic. Each unit had 10,000 lumen light intensity and beam range was up to 800 meters. During the piloting, three alternative light configurations (single blink every 1 s; double blink every 2 s; triple blink every 3 s) were compared to the standardly used reference configuration (without strobe lights), both in the day time and in the night time conditions. The reference configuration had standard train headlights: three continuous white lights, two on the bottom and one on the top. In the alternative configurations, additional blinking LED lights were installed below each of the headlights.	
Test 2 Scope	**Road user role in test**	**LC type in test**
	Car driver	Passive
	Type of test	**Data assessed**
	Online survey based on field-test videos (A3)	Subjective assessments

Coloured road markings on approach to LC

Illustration		
Measure description	The coloured road markings aim to improve the visibility and readability of LC to improve the vigilance of drivers when they approach the LC and to reduce their driving speed.	
Measure Scope	**Target road users**	**Target LC type**
	All road users.	All LCs
	Proposed effect mechanism	**Main psychological function involved**
	Increases awareness of correct behavior / consequences of rule violation	Detection
Measure specification in test	The road marking used in this study included fours stripes: 1) yellow band of 5 cm at 150 meters, 2) orange band of 10 cm at 100 meters, 3) train band 50 cm at 75 meters, and 4) 20 cm red band at 2 meters from the active LC (see illustration).	
Test Scope	**Road user role in test**	**LC type in test**
	Car driver	Active with barrier
	Type of test	**Data assessed**
	Driving simulation (A1)	Driving dynamics, interviews

In-vehicle LC/train proximity warning (1)

Illustration		
Measure description	An application which can be installed on an in-vehicle driver information device or on any common mobile device such as a smartphone or tablet. It warns road users who approach a LC by generating an alert.	
Measure Scope	**Target road users**	**Target LC type**
	Motorized road users	All LCs (most effective at passive)
	Proposed effect mechanism	**Psychological functions**
	Provides up-to-date information about LC status	Rule knowledge, Decision-making
Measure specification in test	The warning consisted of an auditory alert (short train sound) and a visual alert (see illustration) that appeared in a dedicated pop-up window on the screen. Whenever an incoming train was expected to reach the LC within one minute, the warning included an information on the estimated time of train arrival (illustration on right). If no train was approaching or no train information was available, a general LC proximity alert (on left) was given. The application was downloaded and used by more than 600 test vehicles (taxis). The alerts where generated around all 29 LCs in the wider area of Thessaloniki, Greece, within a period of ten months. The time of train arrival could be estimated only for the trains equipped with geolocation devices, which typically operate in the busiest rail line and traverse actively protected LCs with barriers. The alerts where generated at safe distance from the LC. The area around the LC where the alert is triggered was defined in a case approach, considering the topology and surrounding environment of each LC, speed limits, nearby roads and their directions.	
Test Scope	**Road user role in test**	**LC type in test**
	Taxi Drivers	All LCs
	Type of test	Data assessed
	Field test (A4)	Floating car data, backend system data, subjective assessments

2.2.3. Measures for active LCs with barriers

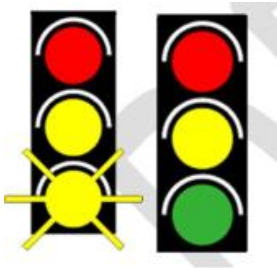
In-vehicle LC/train proximity warning (2)

<table>
<tr><td>Illustration</td><td colspan="2">
</td></tr>
<tr><td>Measure description</td><td colspan="2">This measure aims to improve safety of LCs by supporting the car drivers in adapting their driving speed to the approach of LC by providing different messages to their in-vehicle device.</td></tr>
<tr><td rowspan="4">Measure Scope</td><td>Target road users</td><td>Target LC type</td></tr>
<tr><td>Motorized road users</td><td>All LCs</td></tr>
<tr><td>Proposed effect mechanism</td><td>Psychological functions</td></tr>
<tr><td>Provides up-to-date information about LCstatus</td><td>Rule knowledge, Decision-making</td></tr>
<tr><td>Measure specification in test</td><td colspan="2">The content of these messages varied according to the status of the LC (LC at 300 m, LC closed for construction works, no crossings).</td></tr>
<tr><td rowspan="4">Test Scope</td><td>Road user role in test</td><td>LC type in test</td></tr>
<tr><td>Car driver</td><td>All LCs</td></tr>
<tr><td>Type of test</td><td>Data assessed</td></tr>
<tr><td>Driving simulation (A1)</td><td>Driving dynamics, interviews</td></tr>
</table>

Rings upstream of the LC

Illustration		
Measure description	Installation of two rings on the road upstream of the LC. The lights in the rings start flashing when a train is coming. The rings are targeted to active level crossings with barriers located at rural areas. The second ring must not obscure the visibility of the red flashing light of the LC. Note: Different from a simple "portal", this measure is technologically more challenging	
Measure Scope	**Target road users**	**Target LC type**
	Motorized road users.	Active
	Proposed effect mechanism	**Main psychological function involved**
	Improves LC detection.	Detection
Measure specification in test	The measure consisted of two rings located ahead of the LC: the first one at 150 m, the second at 10 m ahead of the LC. The rings include a set of LEDs and an orange light (diameter of 30 cm, see illustration).	
Test Scope	**Road user role in test**	**LC type in test**
	Car driver	Active with barriers
	Type of test	**Data assessed**
	Driving simulation (A1)	Driving dynamics, interviews

Traffic light

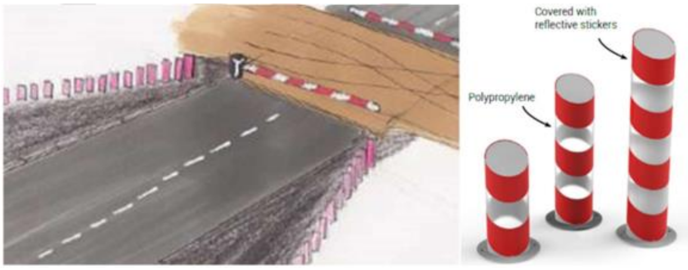
Illustration		
Measure description	Use of traffic lights at LCs instead of LC lights, as road users often respect traffic lights more than LC lights. The traffic lights should be coordinated with an announcement of the LC and with ground loops to manage traffic jam and avoid queuing on the LC. Note on applicability: In LC with barriers, this could be technically coupled to the closing signals. In passive LC it would be effective, too, in case additional detection technology could be installed.	
Measure Scope	**Target road users**	**Target LC type**
	All road users.	Active
	Proposed effect mechanism	**Main psychological function involved**
	Controls the access to / supports the egress from LC	Rule knowledge
Measure specification in test	Two different configurations of this measure were implemented: 1) the flashing red light (R24, France) was replaced with a two-color light (22j) flashing orange, and 2) the flashing red light (R24) was replaced with a green traffic light.	
Test Scope	**Road user role in test**	**LC type in test**
	Car driver	Active with barriers
	Type of test	**Data assessed**
	Driving simulation (A1)	Driving dynamics, interviews

2.2.5. Measures for passive LCs

Blinking amber light with train symbol

Illustration		
Measure description	A blinking amber light with a train symbol is positioned at the side of the road. The blinking train symbol directly ahead of the LC aims to enhance the probability that oncoming trains get detected by increasing road users' awareness that a train might be approaching and thus the motivation to scan the tracks to the left and right.	
Measure Scope	**Target road users**	**Target LC type**
	VRU	Passive
	Proposed effect mechanism	**Main psychological function involved**
	Improves Train/LC Detection	Identification, Rule knowledge
Measure specification in test	The measure was implemented as shown in the illustration, at a passive LC strongly frequented by VRUs. The blinking was activated whenever a road user was detected approaching the LC.	
Test Scope	**Road user role in test**	**LC type in test**
	VRU (mostly bicyclists and pedestrians)	Passive
	Type of test	**Data assessed**
	Field test (A5)	Behavior on approach (kinematics, head movements)

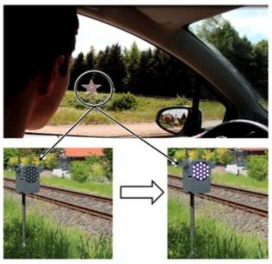
Funnel effect pylons

Illustration		
Measure description	Installation of pylons to create a ‘funnel’ effect (visual impression of narrowing roadway) that motivates drivers to reduce speed on the approach to LCs on straight roads. Note on applicability: conceived for road users traveling at higher speeds, therefore classified MRU, but could also be used for motorcyclists and possibly cyclists	
Measure Scope	**Target road users**	**Target LC type**
	Motorized road users	Passive
	Proposed effect mechanism	**Main psychological function involved**
	Improves LC detection. Increases awareness of correct behaviour / consequences of rule violation	Detection
Measure specification in test	The measure consisted of 15 pylons with a diameter of 20 cm and increasing height, installed upstream of the LC to create the ‘funnel’ effect. The pylons were white and red (see illustration) and covered with reflective stickers.	
Test Scope	**Road user role in test**	**LC type in test**
	Car driver	Active with barriers
	Type of test	**Data assessed**
	Driving simulation (A1)	Driving dynamics, interviews

Message "🡐 Is a train coming? 🡒" written on road

Illustration		
Measure description	The message "🡐 Is a train coming? 🡒" is implemented as a road marking on LC approach. The measure aims at enhancing the probability that road users detect oncoming trains, by providing a reminder of the necessity to scan the tracks for a train to facilitate this target behaviour.	
Measure Scope	**Target road users**	**Target LC type**
	VRU	Passive
	Proposed effect mechanism	**Main psychological function involved**
	Improves Train/LC Detection	Identification, Rule knowledge
Measure specification in test	The text on the message was phrased as a question (🡐 Is a train coming? 🡒), not an instruction, to enhance acceptance and evoke road users' intrinsic interest in knowing whether a train is coming. Two arrows pointing to the left and right were included to illustrate the message and facilitate the allocation of attention to the periphery (see illustration). The message was positioned at 35 m ahead of the LC.	
Test Scope	**Road user role in test**	**LC type in test**
	VRU (mostly bicyclists and pedestrians)	Passive
	Type of test	**Data assessed**
	Field test (A5)	Behavior on approach (kinematics, head movements)

Peripheral blinking lights near the tracks

Illustration		
Measure description	When a car passes an in-road sensor on approach to the LC, two lights located near the tracks to the left and right of the road start blinking. The blinking lights appear in the periphery of the driver's visual field. The salient blinking triggers an automatic and effortless visual orientation reaction in the road user towards the peripheral regions of the level crossing that require visual scanning in order to detect a train (physiological mechanism of exogenous capture of attention; Yantis, 2000) .	
Measure Scope	**Target road users**	**Target LC type**
	Motorized road users.	Passive
	Proposed effect mechanism	**Main psychological function involved**
	Improves train detection	Detection
Measure specification in test	The measure's setup was adapted to the situation at the test LC that allowed free view on the tracks already at 240 m and more ahead of LC: Three posts with blinking lights were implemented at the tracks both to the left and right at 40, 60 and 80 m distance from the road. The blinking was triggered when the ego car was at 250 m ahead of the LC. The blinking sequence started with the two inner lights being activated for 0.1 s, followed by the two lights in between being activated for 0.1 s and, finally, the two outermost lights being activated for 0.1 s. After a pause of 0.1 s with all lights out, the sequence started again, yielding an impression of the blinking "moving" from the center to the periphery on both sides of the road. The blinking continued for 15 s overall.	
Test Scope	**Road user role in test**	**LC type in test**
	Car driver	Passive
	Type of test	**Data assessed**
	Driving simulation (A2)	Gaze, driving dynamics, subjective assessments

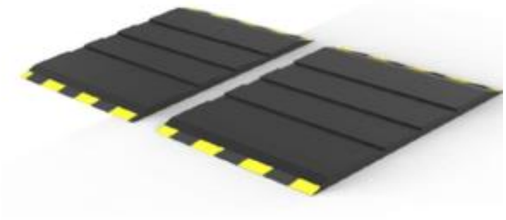
Rumble strips

Illustration		
Measure description	Application of structured or milled markings in the road surface on approach to the LC, to enhance attention and induce speed reduction in drivers.	
Measure Scope	**Target road users**	**Target LC type**
	Motorized road users	Passive
	Proposed effect mechanism	**Psychological functions**
	Reduces approach speed of vehicles.	Identification;Decision-making; execution
Measure specification in test	Rumble strips occured two times on LC approach: The first implementation started at 300 m distance to the LC, coinciding with the sight of the three-striped post (German sign 157, announcing LC, positioned at 240m; see illustration on the right), the second one at 90 m distance to the LC, coinciding with the one-striped post (German sign 162, announcing LC, positioned at 80 m).	
Test Scope	**Road user role in test**	**LC type in test**
	Car driver	Passive
	Type of test	**Data assessed**
	Driving simulation (A2)	Gaze, driving dynamics, subjective assessments

Sign “← Is a train coming? →”

Illustration		
Measure description	Installation of a sign „Is a train coming“ on approach to the LC. The measure aims at enhancing the probability that road users detect oncoming trains, by providing a reminder of the necessity to scan the tracks for a train to facilitate this target behaviour.	
Measure Scope	**Target road users**	**Target LC type**
	Motorized road users	Passive
	Proposed effect mechanism	**Main psychological function involved**
	Improves Train/LC Detection	Detection
Measure specification in test	The text on the sign was phrased as a question (← Is a train coming? →), not an instruction, to enhance acceptance and evoke road users’ intrinsic interest in knowing whether a train is coming. Two arrows pointing to the left and right and two train pictograms where the arrows are pointing were included to illustrate the message and facilitate the allocation of attention to the periphery. Moreover, the pictograms contribute to making the sign comprehensible also to road users who cannot read or do not understand the given language. The sign was positioned at 100 m ahead of the LC (see illustration).	
Test Scope	**Road user role in test**	**LC type in test**
	Car driver	Passive
	Type of test	**Data assessed**
	Driving simulation (A2)	Gaze, driving dynamics, subjective assessments

Speed bumps and flashing posts

Illustration	 	
Measure description	Installation of speed bumps, combined with flashing posts on approach to the LC. The measure aims to improve the visibility and detectability of the LC in order to improve the vigilance of drivers as they approach the LC and induce speed reduction.	
Measure Scope	**Target road users**	**Target LC type**
	Motorized road users	Active
	Proposed effect mechanism	**Main psychological function involved**
	Reduces approach speed of vehicles Improves LC detection	Decision-making; Detection
Measure specification in test	The posts were equipped with a red LED lamp (see illustration). Three poles working in alternating flicker were located at 150, 100 and 50 m from the LC on the right edge of the roadway (Figure 18). The bumps were located at 150, 100 and 50 m from the LC. The number of inner lines differed according to their location on LC approach (1, 2 or 3 lines).	
Test Scope	**Road user role in test**	**LC type in test**
	Car driver	Active with barriers
	Type of test	**Data assessed**
	Driving simulation (A2)	Driving dynamics, interviews

2.4. Procedure and analysis

In the course of the project, the HFAT was filled in two times by all test site leaders whose pilot involved the assessment of data on road user behavior and experience connected with piloted safety measures. The first assessment was done in December 2018 and January 2019. This was called the “baseline” assessment as one of its aims was to assess baseline data (i.e. behavioral data of LC users as observed without any additional safety measures applied) in the large field test at Thessaloniki. However, there were two more objectives associated with this assessment. First, the available evidence from the literature concerning the measure (or similar measures) was to be collected. This was done for the field tests where baseline data was already collected at that time, as well as for the simulation tests in which the baseline or control condition was included as one of multiple conditions within one test session and the studies were scheduled for spring and summer 2019. Second, the test site leaders were to practice the application of the HFAT and give feedback on its suitability and usability. The feedback was used to elaborate the tool for the second assessment. The most important change as regards content concerned the section *behavioral safety* (green HFAT section), where the first version of the HFAT made a distinction in the early stages of information processing between *detection* and *identification*, analogous to the model of human information processing (Grippenkoven and Dietsch, 2015; Havârneanu et al., 2018). In the review phase after the first assessment it was decided to integrate the two stages into one, due to the observation from the first application that measures that help detection mostly also help identification, and behavioral indicators as they can be collected in applied research (e.g. gaze metrics, answers to the question “did you see the …?”) mostly make it hard to clearly assign the results to just one of the two stages.

The second assessment, called “test” assessment, was done by each test site leader by the time that data analyses from the pilot were available to show how LC users reacted to the piloted measures. The results presented in this deliverable are based on the HFAT inputs from this second assessment for all participating test sites. These inputs also contained the results from the baseline assessment. The HFAT inputs were processed with the aim of integrating the results for all measures from a given HFAT section, in order to allow their joint consideration and comparison across the measures. To this aim, all the information obtained on applicability (orange HFAT section), behavioral safety effects (green), and acceptability (blue) was visualized in tables to enable an easy overview. To enhance readability in the following results part, all results tables are introduced first (sections 3.1 to 3.3). The results are then summarized (section 3.4), before a concluding overall assessment of the measures is made (section 3.5).

3. RESULTS OF THE HUMAN-FACTORS ASSESSMENT

3.1. Applicability

The first HFAT section (orange) contains classification criteria to describe the scope of the safety measure under assessment, i.e. to define in detail the context and environment in which the measure is expected to be effective. There are five subsets of characteristics related to this description:

- *Applicability to different LCs,*
- *feasibility under different environmental conditions,*
- *applicability to different types of users,*
- *adaptation to individual characteristics & conditions of users*, and
- *intended effect mechanism*.

In most of these subsets, there are further sub categories defined. The first set for example is split up further into *types of LCs* and *characteristics of LCs*, of which the first one deals with the existing type of protection applied at a given LC, and the second with other characteristics of the infrastructure and traffic environment, such as traffic volume, road quality, availability of electricity, or crossing angle.

In the following, the categorizations attained for the 13 piloted measures are presented in separate tables for each of the five subsections. The measures appear in the table rows in the same order as introduced before (cf. section 2.2), which is also maintained in all following results sections.

Table 3. Applicability to different LCs.

Measure	Sub category															
	Types of LCs								Characteristics of LCs							
	Passive LCs without any warning device	Active (manual)	Active LCs with half barriers	Active LCs with full barriers	Active LCs with skirts for pedestrians	Active LCs with light and sound warning	Active LCs with other warning devices	Active LCs with traffic lights	LCs with low vehicle traffic	LCs with high vehicle traffic	LCs with paved road	LCs with gravel road	LCs with availability of electricity	LCs with low usage / not used at all	LCs with sharp / wide crossing angle	Other (specify)
Blinking lights for locomotive front	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Coloured road markings on approach to LC	X	X	X	X		X	X	X	X	X	X		X	X	X	
In-vehicle proximity warning (1)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
In-vehicle proximity warning (2)		X	X	X		X	X	X	X	X	X	X	X	X	X	
Rings upstream of the LC		X	X	X		X	X	X	X	X	X	X	X	X	X	
Traffic light		X	X	X		X	X		X	X	X	X	X	X	X	
Blinking amber light with train symbol	X								X	X	X	X	X	X	X	
Funnel effect pylons	X	X	X	X		X	X	X	X	X	X	X	X	X	X	
Message "Is a train coming?" on road	X								X	X	X		X	X	X	
Peripheral blinking lights	X								X	X	X	X	X	X	X	
Rumble strips	X					X			X	X	X		X	X	X	
Sign „Is a train coming“	X								X	X	X	X	X	X	X	
Speed bumps and flashing posts		X	X	X		X	X	X	X	X	X		X	X	X	

Table 4. Feasibility under different environmental conditions.

Measure	Sub category											
	Time of the day					Weather conditions					Setting	
	Daylight	Darkness	Dusk	Dawn	Peak traffic hours	Rain	Snowfall	Slipperiness	Fog	Bright sun / glare	urban	rural
Blinking lights for locomotive front	X	X	X	X	X	X	X	X	X		X	X
Coloured road markings on approach to LC	X		X	X	X	X			X	X	X	X
In-vehicle proximity warning (1)	X	X	X	X	X	X	X	X	X	X	X	X
In-vehicle proximity warning (2)	X	X	X	X	X	X	X	X	X	X	X	X
Rings upstream of the LC	X	X	X	X	X	X	X	X	X	X	X	X
Traffic light	X	X	X	X	X	X	X	X	X	X	X	X
Blinking amber light with train symbol	X	X	X	X	X	X	X	X		X	X	X
Funnel effect pylons	X	X	X	X	X	X	X	X	X	X	X	X
Message "Is a train coming?" on road	X		X	X	X	X		X		X	X	X
Peripheral blinking lights	X	X	X	X	X	X	X	X		X	X	X
Rumble strips	X	X	X	X	X	X		X	X	X	X	X
Sign „Is a train coming“	X	X	X	X	X	X	X	X			X	X
Speed bumps and flashing posts	X	X	X	X	X	X	X	X	X	X	X	X

Table 5. Applicability to different types of users.

Measure	Sub category									
		MRU						VRU		
	all road users	cars	motorbikes / mopeds	trucks / heavy vehicles	buses / coaches	agricultural vehicles	Other (specify)	pedestrians	cyclists	other (specify)
Blinking lights for locomotive front	X	X	X	X	X	X	X	X	X	X
Coloured road markings on approach to LC		X	X	X	X	X			X	
In-vehicle proximity warning (1)		X		X	X		X			X
In-vehicle proximity warning (2)		X		X	X	X				
Rings upstream of the LC		X	X	X	X	X			X	
Traffic light		X	X	X	X	X			X	
Blinking amber light with train symbol	X	X	X	X	X	X		X	X	
Funnel effect pylons		X	X	X	X	X		X	X	
Message "Is a train coming?" on road	X	X	X	X	X	X		X	X	
Peripheral blinking lights	X	X	X	X	X	X	X	X	X	X
Rumble strips		X		X	X					
Sign „Is a train coming“	X	X	X	X	X	X	X	X	X	X
Speed bumps and flashing posts		X	X	X	X	X			X	

Table 6. Adaptation to individual characteristics and conditions of users.

Measure	Sub category															
	Gender		Age			Disability					Under influence of			Under skill impairing states		
	Male	Female	Children	elderly	All ages	vision loss and blindness	hearing loss and deafness	intellectual disability	reduced mobility	other (specify)	alcohol	drugs	medication	fatigue	stress	risk-seeking personality
Blinking lights for locomotive front	X	X	X	X	X		X	X	X		X	X	X	X	X	X
Coloured road markings on approach to LC	X	X	X	X	X											
In-vehicle proximity warning (1)	X	X		X	X		X		X		X	X	X	X	X	
In-vehicle proximity warning (2)				X	X											
Rings upstream of the LC	X	X	X	X	X											
Traffic light	X	X	X	X	X											
Blinking amber light with train symbol	X	X	X	X	X		X	X	X		X	X	X	X	X	X
Funnel effect pylons	X	X	X	X	X											
Message "Is a train coming?" on road	X	X	X	X	X		X		X		X	X	X	X	X	X
Peripheral blinking lights	X	X	X	X	X		X	X	X		X	X	X	X	X	X
Rumble strips	X	X	X	X	X		X		X		X	X	X	X	X	X
Sign „Is a train coming“	X	X	X	X	X		X		X		X	X	X	X	X	X
Speed bumps and flashing posts	X	X	X	X	X											

Table 7. Intended effect mechanism.

Measure	Improves the detection of train	Improves the detection of LC	Controls access to / supports egress from the LC	Reduces the approach speeds of vehicles	Increases the user's awareness of correct behaviour and consequences of rule violation	Improves the physical environment of LC	Improves the possibilities of VRU to cross LC safely	Provides up-to-date information about LC status	Supports the LC safety actions	Other (specify)
Blinking lights for locomotive front	X				X			X		
Coloured road markings on approach to LC		X		X	X					
In-vehicle proximity warning (1)	X	X						X		
In-vehicle proximity warning (2)		X			X					
Rings upstream of the LC		X		X	X					
Traffic light		X			X					
Blinking amber light with train symbol		X			X					
Funnel effect pylons		X		X	X					
Message "Is a train coming?" on road		X			X					
Peripheral blinking lights	X	X			X					
Rumble strips		X		X	X					
Sign „Is a train coming"	X			X	X					
Speed bumps and flashing posts		X		X	X					

3.2. Behavioral safety effects

The second HFAT section (green) is used to assess the effects of LC safety measures on road user behaviour. The effects are assessed with respect to four stages of information processing at which the measure could in principle exert an influence:

- *Detection and identification,*
- *rule knowledge,*
- *decision-making*, and
- *behavioral execution.*

The completion of the HFAT concerning the effects of safety measures at each of these stages comprises two steps. In the first step, relevant empirical evidence is collected that provides information on how the measure (or a theoretically similar measure) affects road user behavior at this stage. In the second step, a numerical score is assigned based on the available evidence and / or theoretical assumptions, to express to what extent the measure is expected to facilitate safe road user behavior at a given stage.

For the collection of evidence, there are eight separate categories in which the available findings are sorted. These eight categories result from the combinations of three dimensions. The most important one refers to the condition under which the respective finding was obtained: *base* – the finding describes how road users behave without the measure under otherwise similar conditions – vs. *test* – the finding describes how road users behave when the measure is applied. The second important dimension is the time scale of the observation: To take account of the fact that behavioral changes can be modified over time by behavioral adaption, the available findings are to be classified as *short-* or *long-term* effects. Finally, to highlight the new evidence yielded in the course of the SAFER-LC project, a distinction is made between results found in the *literature* and findings obtained in the *pilot.*

Based on the evidence collected for each stage of information processing, the reviser is asked to assign a score to express to what extent the measure will likely facilitate safe road user behavior at that stage. The score can take values between *0* (the measure does not influence information processing at the given stage) and *5* (the measure maximally facilitates information processing at the given stage towards the desired outcome). The value *N* is assigned in case the measure must be considered as even impeding the desired processes. In some cases, no score was assigned at a given stage. In these cases, the value *NA* is shown in the table.

In the following, the results obtained for the 13 measures are presented in separate tables for each of the four stages of information processing. Each table juxtaposes the scores assigned to each measure with the evidence these scores are based on. To enable an overview, the scores are colour-coded, with darker shading indicating higher values. The evidence is represented in an aggregated form: If there is evidence available in one of the eight categories, the respective cell is shaded and a short reference appears in it (e.g. *Ab1)*. The reference consists of a capital letter indicating one of the measures (e.g. *A - blinking lights for locomotive front, B - coloured road markings*, etc*.),* followed by a lowercase letter (either *b* or *t,* for findings from the baseline vs. test condition), and a consecutive number (indicating each unique combination of information processing stage, time scale and source of the finding). All the findings referred to in a given overview table are listed below it in a subsequent evidence table. Each entry in the evidence table starts with the reference from the overview, followed by the respective finding. The source of the finding is indicated as either a reference to the literature (short methods summaries of all studies cited are provided in Table 8) or a reference to the annex describing the pilot in which the finding was obtained (A1 to A5, see Table 2 and Annexes). As an example, to compare the results obtained for measure A, the *blinking lights for locomotive front*, in the baseline condition vs. the test condition in the pilot study, the reader needs to consult the entries *Ab3* and *At4* in the evidence table.

Table 8. Short methods summaries of the studies cited in the collection of evidence from the literature.

Citation	Publication title	Methods
Cairney (2003)	Prospects for improving the conspicuity of trains at passive railway grade crossings	Literature review on effects of additional lights on trains (7 studies)
Carroll, Multer & Markos (1995)	Safety of Highway-Railroad Grade Crossings: Use of Auxiliary External Alerting Devices to Improve Locomotive Conspicuity	Study. Data: detectability of train, Setting: Controlled field test. Three variants of supplemental light systems for trains tested: crossing light (blinking headlights) – ditch light (intense static headlight) – strobe lights (intense source of light pulsating at high frequency). Subjects were seated on a chair at a certain distance from the LC and were aware that a train would approach at some point in time. In addition to the train detection task, a computer-based monitoring task was administered to simulate the attentional demands in a traffic situation.
Grippenkoven & Dietsch (2015)	Gaze direction and driving behavior of drivers at level crossings	Study. Data: Driving data, gaze data, Setting: driving study in real traffic, approach of two level crossings (one with passive protection, the other with active light-signal protection). Suburban area with maximum speed 50 km/h; participants unfamiliar with the two LC and unaware of study purpose
Grippenkoven, Thomas & Lemmer (2016)	PeriLight – effektive Blicklenkung am Bahnübergang	Study. Data: driving data, gaze data, Setting: driving study in real traffic in daylight and at night, two encounters with the same LC (blinking peripheral lights inactive vs. active), participants unaware of LC focus.
Hore-Lacy (2008)	Rumble strip effectiveness at rural intersections and railway level crossings	Study. Data: Speed on approach to intersection, driving around rumble strips, Setting: Pilot in real traffic (before-after with control sites) with rumble strips installed at 4 passive LCs
Liu, Bartnik, Richards & Khattak (2016)	Driver behavior at highway–rail grade crossings with passive traffic controls: A driving simulator study	Study. Data: Looking behavior (observed head movements), speed on LC approach, stopping. Setting: Simulator study of driver behaviors at passive LC equipped with different traffic signs (St. Andrew's Cross, Stop, Yield)
Noyce & Fambro (1998)	Enhanced traffic control devices at passive highway-railroad grade crossings	Study. Data: speed reduction on approach to LC, driver survey on conspicuity and comprehension of strobe light with sign "Look for train", observed driving behaviour at night time (vehicle manoeuvres as braking, weaving, swerving); Setting: Pilot in real-traffic (before-after) with sign "Look for train" at crossing (black on yellow) and double arrow, supplemented with vehicle-activated strobe light

Citation	Publication title	Methods
Radalj & Kidd (2005)	A Trial with Rumble Strips as a Means of Alerting Drivers to Hazards at Approaches to Passively Protected Railway Level Crossings on High Speed Western Australian Rural Roads	Study. Data: Speed on approach to LC. Setting: Pilot test in real traffic (before-after) with rumble strips installed at 11 passive LCs with Give-Way signs (one group of rumble strips) and 3 passive LCs with Stop signs (four groups of rumble strips)
Shinar & Raz (2007)	Driver response to different railroad crossing protection systems	Study. Data: speed at 200 and 80 m before the tracks, stopping. Setting: Field observation in rural area of drivers' approach to the same LC in five conditions: passive, active with flashing lights (inactive vs. active), active protection with flashing light and half-barriers (inactive, active). Drivers unaware of being observed. Visibility of approaching trains limited for drivers because of trees, therefore full stop or speed reduction to at most 10 km/h necessary to detect an oncoming train in time to stop before the tracks.
Skládaný et al (2016)	Entwicklung von Rüttelstreifen zur Vermeidung von Fehlverhalten an Eisenbahnkreuzungen: Ergebnisbericht zum Forschungsprojekt „RÜTTLEX". (Engl.: Development of rumble strips to avoid maladaptive behavior at LCs: Report on the results of the research project RÜTTLEX)	Study. Data: Speed on approach to intersection, lane choice (driving around rumble strips), driver survey (detection of test object [toy parrot] that was positioned at the tracks; qualitative assessment of LC); Setting: Pilot in real traffic (before – after [short] – after [long – 1 year]) with rumble strips installed at a passive LC

3.2.2. Detection and identfication

Table 9. Overview of the HFAT assessment of the safety effects of the piloted measures at the stage of detection and identification.

	Measure	Score	Evidence				
			Timescale	Literature		Pilot	
				base	test	base	test
A	Blinking lights for locomotive front	5	short-term	Ab1	At2	Ab3	At4
			long-term				
B	Coloured road markings on approach to LC	3	short-term				Bt4
			long-term				
C	In-vehicle proximity warning (1)	5	short-term				Ct4
			long-term			Cb7	
D	In-vehicle proximity warning (2)	5	short-term				Dt4
			long-term				
E	Rings upstream of the LC	3	short-term				Et4
			long-term				
F	Traffic light	4	short-term				Ft4
			long-term				
G	Blinking amber light with train symbol	3	short-term				Gt4
			long-term				
H	Funnel effect pylons	0	short-term				Ht4
			long-term				
I	Message "Is a train coming?" on road	1	short-term				It4
			long-term				
J	Peripheral blinking lights	4	short-term	Jb1	Jt2	Jb3	Jt4
			long-term				
K	Rumble strips	2	short-term	Kb1	Kt2	Kb3	Kt4
			long-term	Kb5	Kt6		
L	Sign Look for train	3	short-term		Lt2	Lb3	Lt4
			long-term				
M	Speed bumps and flashing posts	4	short-term				Mt4
			long-term				

Table 10. Evidence collected on the measures' effects on detection and identification.

Ref.	Context	Findings
Ab1	base-literature-short	Baseline (train with normal lighting) was detected worse than experimental conditions (Carrol, Multer, & Markos, 1995).
At2	test-literature-short	trains were detected at greater distance from LC with all three of the tested systems. Crossing light judged most effective (Carrol, Multer, & Markos, 1995).

Ref.	Context	Findings
Ab3	base-pilot-short	baseline without train: Proportion of participants who scanned periphery for a train and mean number of periphery fixations (SD) between 140 and 40 m ahead of LC: left side: 65% and 1.2 (1.2), right side: 46% and 0.88 (1.3). - baseline with train (normal train): All but one participant detected the train and let it pass before they crossed the LC; distance driven before first fixation on train after it became visible *M* = 30.5 m (*SD* = 27.6); *Max* = 110 m (A2)
At4	test-pilot-short	All train conditions: Proportion of participants who scanned periphery and mean number of periphery fixations (SD) between 140 and 40 m ahead of LC: left side: 75.6% and 2.1 (2.4), right side: 53.7% and 1.3 (1.8); train had already been detected before except in one case. Blinking train: All participants detected the train and let it pass before they crossed the LC; distance driven before first fixation on train after it became visible *M* = 8.2 m (*SD* = 9.3); *Max* = 37 m (A2). In the questionnaire, improved detectability of the train was mentioned often as a benefit of blinking lights (A3).
Bt4	test-pilot-short	Of 27 subjects interviewed, 20 reacted (by braking of ceasing to accelerate) to the sight of the marking on the ground. To note that only 2 subjects saw the marking TRAIN and the red line, the others saw the white marking train alone. This reaction translates into a reduction of speed at the approach of the LC (A1).
Ct4	test-pilot-short	Phase 2 Questionnaire relevant questions: 7. How easy is it to detect a LC and approaching trains using the in-car alert system? 8.How easy is it to identify LCs that you were not previously aware of or a possible danger at a LC using the in-car alert system?. Q7: 14.8% and 3.7% answered 'Not at all' and 'Slightly' respectively. Almost 29.63% of drivers stated that they find it 'completely' easy. Q8: 11.11% and 7.41% answered 'Not at all' and 'Slightly' respectively, while 41.7% of drivers stated that they find it 'completely' easy (A4).
Cb7	base-pilot-long	Phase 1 Questionnaire relevant questions:. 7.How easy is it to detect the presence of a LC or an approaching train based on the existing LC safety measures (e.g. signs)? 8.How easy is it to identify LCs that you were not previously aware of or a possible danger at a LC based on the existing LC safety measures (e.g. signs)? Q7: Approximately 10% and 26% answered 'Not at all' and 'Slightly' respectively. Less than 10% of drivers stated that they find it 'completely' easy. Similar results for Q8: 9.3% and 27.3% answered 'Not at all' and 'Slightly' respectively, while less than 10% of drivers stated that they find it 'completely' easy (A4).
Dt4	test-pilot-short	We had 25 subjects for the different connected situations, but 2 subjects were sick and could not do all the situations. 70% of the subjects reacted to the situations with a message associated with a beep giving information of danger (LC closed, LC in work or LC out of order) and state it allows them to anticipate their speed upstream of the LC and to better prepare for the stop. For situations where the message is not received, the subjects will resume their standard behavior and will not be disturbed about receiving nothing (A1).
Et4	test-pilot-short	Of 29 subjects, more than half react to the sight of the arch, the others remain in a standard approach. Half of them (i.e. a quarter overall) perceived the fires on the ark (A1).
Ft4	test-pilot-short	For the situation with the orange traffic light: the subjects slow down at the sight of the flashing light. However, not all of them relate it to the crossing and some are looking for a road intersection. For the situation with the green traffic light: the subjects do not slow down at all, some even accelerate (A1).
Gt4	test-pilot-short	As increased visual search behaviour was observed with the measure (cf. Gb11 and Gt12), it also increases the probability to detect an oncoming train (A5).
Ht4	test-pilot-short	Of 28 subjects, less than 10% react to the sight of the measure by decelerating on approach to the LC, 60% didn't see the measure, the others saw it but didn't understand the measure (A1).
It4	test-pilot-short	Increased visual search behaviour was not observed with the measure in the sample of mainly bicyclist, but is expected in pedestrians (cf. Ib11 and It12), increasing the probability to detect an oncoming train (A5).
Jb1	base-literature-short	Looking behavior (head movement) frequency with no train coming: 71%, with train coming: 72% (Liu, Bartnik, Richards, & Khattak, 2016). Only 33% of participants (8 of 24) checked the tracks for a train before crossing, while all participants fixated at least once on the LC signs (not: tracks); i.e. were aware of the LC (Grippenkoven & Dietsch, 2015). Daylight - visual check of tracks to left: 26% (8 of 30) of participants, right: 50% (15 of 30). Night - left: 29% (5 of 17), right: 35% (6 of 17) (Grippenkoven et al., 2016).
Jt2	test-literature-short	Daylight - looking left: 73% (22 of 30), right: 73% (22 of 30). Night - looking left: 88% (15 of 17), right: 88% (15 of 17) (Liu, Bartnik, Richards, Khattak, 2016).

Ref.	Context	Findings
Jb3	base-pilot-short	Proportion of participants who scanned periphery for a train and mean number of periphery fixations (*SD*) between 140 and 40 m ahead of LC: left side: 65% and 1.2 (1.2), right side: 46% and 0.88 (1.3) (A2)
Jt4	test-pilot-short	Proportion of participants who scanned periphery for a train and mean number of periphery fixations (*SD*) between 140 and 40 m ahead of LC: left side: 84% and 3.2 (3.1), right side: 65% and 2.5 (3.7) (A2)
Kb1	base-literature-short	22.7% and 39.2% of drivers interviewed (driving in direction 1 and 2) noticed the test object at the tracks (i.e. had obviously looked for a train) (Skládaný, Tučka et al., 2016).
Kt2	test-literature-short	61.3 % and 18.8% of drivers interviewed (driving in direction 1 and 2) noticed the test object at the tracks (drop to 18% interpreted as “drivers checked tracks at an earlier point at which the test object could not yet be detected”) (Skládaný, Tučka et al., 2016).
Kb3	base-pilot-short	Proportion of participants who scanned periphery for a train and mean number of periphery fixations (*SD*) between 140 and 40 m ahead of LC: left side: 65% and 1.2 (1.2), right side: 46% and 0.9 (1.3) (A2).
Kt4	test-pilot-short	Proportion of participants who scanned periphery for a train and mean number of periphery fixations (*SD*) between 140 and 40 m ahead of LC: left side: 67% and 1.7 (1.5), right side: 40% and 0.9 (1.4) (A2).
Kb5	base-pilot-long	video observation: no driver drove on the opposite lane, 3 of 77 (4%) drove in the middle of the road (Skládaný, Tučka et al., 2016).
Kt6	test-literature-long	video observation: 3 of 84 drivers (3%) drove on the opposite lane, 4 drivers (5%) drove in the middle of the road, 35% avoided the RSs with the wheels on the left or only passed them with part of the left wheel width; 57% passed the RSs with all wheels (no avoidance tendency) (Skládaný, Tučka et al., 2016).
Lt2	test-literature-short	18 of 33 drivers surveyed indicated they observed the sign. 11 of the 18 who reported observing the sign could recall the wording (exact or similar) (Noyce & Fambro, 1998).
Lb3	base-pilot-short	Proportion of participants who scanned periphery for a train and mean number of periphery fixations (SD) between 140 and 40 m ahead of LC: left side: 65% and 1.2 (1.2), right side: 46% and 0.88 (1.3) (A2).
Lt4	test-pilot-short	Proportion of participants who scanned periphery for a train and mean number of periphery fixations (SD) between 140 and 40 m ahead of LC: left side: 80% and 2.4 (1.9), right side: 47% and 1.22 (1.8) (A2).
Mt4	test-pilot-short	Of 28 subjects, more than half react to the bumps by decelerating on approach to the LC (A1).

3.2.4. Rule-knowledge

Table 11. Overview of the HFAT assessment of the safety effects of the piloted measures at the stage of the activation of rule-knowledge.

	Measure	Score	Evidence				
			Timescale	Literature		Pilot	
				base	test	base	test
A	Blinking lights for locomotive front	4	short-term	Ab9	At10		At12
			long-term				
B	Coloured road markings on approach to LC	3	short-term				Bt12
			long-term				
C	In-vehicle proximity warning (1)	1	short-term				Ct12
			long-term			Cb15	
D	In-vehicle proximity warning (2)	4	short-term				Dt12
			long-term				
E	Rings upstream of the LC	2	short-term				Et12
			long-term				
F	Traffic light	3	short-term				Ft12
			long-term				
G	Blinking amber light with train symbol	3	short-term			Gb11	Gt12
			long-term				
H	Funnel effect pylons	0	short-term				Ht12
			long-term				
I	Message "Is a train coming?” on road	2	short-term			Ib11	It12
			long-term				
J	Peripheral blinking lights	4	short-term	Jb9	Jt10		Jt12
			long-term				
K	Rumble strips	2	short-term	Kb9	Kt10		Kt12
			long-term				
L	Sign Look for train	4	short-term		Lt10		Lt12
			long-term				
M	Speed bumps and flashing posts	3	short-term				Mt12
			long-term				

Table 12. Evidence collected on the measures’ effects on the activation of rule-knowledge.

Ref.	Context	Findings
Ab9	base-literature-short	Less accurate judgement of time to arrival with normal train lighting reported in two studies (Cairney, 2003).
At10	test-literature-short	More accurate judgement of time to arrival with additional lights reported in two studies (Cairney, 2003).
At12	test-pilot-short	92% of participants reported they understood what the blinking train meant. High mean value on subjective rating “easy to understand”: M = 4.7, SD = 1.7 (scale: 1-not at all to 6-completely true) (A2).

Ref.	Context	Findings
Bt12	test-pilot-short	Heterogenous results in terms of understanding the measure's meaning: To some participants, having this marking on the ground is a repetition of the A7 panel and is perceived as a duplicate. Others were irritated by the rendering in the simulator, some thought to see a bump, some focused on reading the text and were surprised to encounter the LC so quickly, others interpreted the measure as a STOP and finally others perceived the arrow as encouraging acceleration (A1).
Ct12	test-pilot-short	Phase 2 Questionnaire relevant question: 9.To what extent does the in-car alert system help you to know how to cross LCs safely in Thessaloniki? Q9: 28.5% of respondents answered 'moderately', 50% 'considerably' or 'completely' (A4).
Cb15	base-pilot-long	Phase 1 Questionnaire relevant question: 9.To what extent do the current safety measures at LCs in Thessaloniki help you to know how to cross safely? Q9: more than 41% of respondents answered 'slightly' or 'not at all'. Another 40.5% 'moderately' and only 18% 'considerably' or 'completely' (A4).
Dt12	test-pilot-short	The majority understands that messages are sent to anticipate situations upstream of the LC. However, some subjects state they prefer to concentrate on their behavior because receiving messages on a screen distracts them and forces them to take their eyes off the road, which is dangerous for them (A1).
Et12	test-pilot-short	Of the 29 subjects, 15 respond to the sight of the ark by imagining for 9 of them a danger and / or the arrival of one and five subjects who question the meaning of the ark. The others remain in a standard approach. 69% of the subjects interpreted this measure as a decoration of village entrance and did not directly relate it to the LC. Some subjects were so focused on the measure that they missed some information such as the A7 panel or the LC view and may have been surprised at the sight of the LC closing. Among the subjects who took the ark for a decoration, 20% associated it with the LC when seeing the panel A7. Only 17% of the subjects understand that arches with fires announce the closure of the LC. 14% of the subjects state not to understand the measure. The subjects are often disturbed by the rings and do not necessarily relate them to the LC (A1).
Ft12	test-pilot-short	The green light is interpreted as a "GO" and that there is no hazard while the orange light is interpreted as "attention, hazard" so the subjects will decelerate on approach (A1).
Gb11	base-pilot-short	Proportion of participants who scanned periphery for a train ahead of LC: left side: 88%; right side: 90%; both ways: 83% neither way: 4% (A5).
Gt12	test-pilot-short	Proportion of participants who scanned periphery for a train ahead of LC: left side: 95%; right side: 94%; both ways: 90%, neither way 1% (A5).
Ht12	test-pilot-short	Only 10% of subjects understand that there is a hazard zone associated with the funnel effect measure and reduce speed (A1).
Ib11	base-pilot-short	Proportion of VRUs observed who turned their head in a given direction ahead of LC: left side: 88%; right side: 90%; both ways: 83%; neither way: 4% (A5).
It12	test-pilot-short	Proportion of VRUs observed who turned their head in a given direction ahead of LC: left side: 86%; right side: 93%; both ways: 84%, neither way 5%. Note: The majority of VRUs were bicyclists. The measure is expected to be more effective with pedestrians as they have more time to read and process the message. The message explicitly addresses the crucial rule-knowledge (A5).
Jb9	base-literature-short	Only 33% of participants (8 of 24) checked the tracks for a train before crossing (Grippenkoven & Dietsch, 2015). Daylight - looking left: 26% (8 of 30) of participants, right: 50% (15 of 30), Night - looking left: 29% (5 of 17), right: 35% (6 of 17) (Grippenkoven et al., 2016) drivers were about 8 km/h slower at 80 m (48 km/h) in front of the LC than at 200 m (56 km/h). 30% of drivers exceeded the 55km/h per hour needed at most to be able to stop before the tracks. No significant difference in speed between active and passive protection (Shinar & Raz, 2007).
Jt10	test-literature-short	Daylight - looking left: 73% (22 of 30), right: 73% (22 of 30). Night - looking left: 88% (15 of 17), right: 88% (15 of 17) (Grippenkoven et al., 2016).
Jt12	test-pilot-short	55% of participants reported they understood what the blinking peripheral lights meant. Moderate mean value on subjective rating "easy to understand": M = 3.9, SD = 1.64 (scale: 1-not at all to 6-completely true) (A2).

Ref.	Context	Findings
Kb9	base-literature-short	see Ib1
Kt10	test-literature-short	see It2
Kt12	test-pilot-short	31% of participants reported they understood what the rumble strips meant. Moderate mean value on subjective rating "easy to understand": M=3.4, SD=1.82 (scale: 1-not at all to 6-completely true) (A2).
Lt10	test-literature-short	9 of the 11 drivers who recalled the wording of the sign indicated the sign caused them to drive with additional caution because they assumed that transportation officials would not have installed it if the crossing was not unsafe (Noyce & Fambro, 1998).
Lt12 -	test-pilot-short	94% of participants reported they understood what the sign meant. High mean value on subjective rating "easy to understand": M =5.2, SD = 1.05 (scale: 1-not at all to 6-completely true) (A2).
Mt12	test-pilot-short	Subjects understand that the bumps announce a hazard, but few relate them to the LC. Very few subjects noticed the side light beacons because most focused on the bumps (A1).

3.2.6. Decision-making

Table 13. Overview of the HFAT assessment of the safety effects of the piloted measures at the stage of decision-making.

	Measure	Score	Timescale	Evidence: Literature base	Evidence: Literature test	Evidence: Pilot base	Evidence: Pilot test
A	Blinking lights for locomotive front	4	short-term	Ab17	At18	Ab19	At20
			long-term				
B	Coloured road markings on approach to LC	NA	short-term				
			long-term				
C	In-vehicle proximity warning (1)	4	short-term				Ct20
			long-term			Cb23	
D	In-vehicle proximity warning (2)	NA	short-term				
			long-term				
E	Rings upstream of the LC	NA	short-term				
			long-term				
F	Traffic light	NA	short-term				
			long-term				
G	Blinking amber light with train symbol	2	short-term			Gb19	Gt20
			long-term				
H	Funnel effect pylons	NA	short-term				
			long-term				
I	Message "Is a train coming?" on road	1	short-term			Ib19	It20
			long-term				
J	Peripheral blinking lights	4	short-term	Jb17		Jb19	Jt20
			long-term				
K	Rumble strips	2	short-term	Kb17	Kt18	Kb19	Kt20
			long-term		Kt22		
L	Sign Look for train	4	short-term		Lt18	Lb19	Lt20
			long-term				
M	Speed bumps and flashing posts	NA	short-term				
			long-term				

Table 14. Evidence collected on the measures' effects on decision-making.

Ref.	Context	Findings
Ab17	base-literature-short	see Ab9
At18	test-literature-short	see At10
Ab19	base-pilot-short	Mean speed [km/h] with (SD) on LC approach (without train) at:. 240 m: 53.0 (6.1). - 160 m: 51.9 (6.4). - 80 m: 50.1 (6.2). - 40 m: 47.7 (8.1) - 30 m: 47.0, (9.2). - 20 m: 45.5 (12.3). - 10 m: 46.0 (11.6). - 0 m: 47.2 (9.8). - Allowed max. speed was 50 km/h (A2).

Ref.	Context	Findings
At20	test-pilot-short	Mean speed profile on approach to LC: Speed reduction starting at around 240 m ahead of the LC (net 30 m earlier compared to normal train) and reaching its maximum of 16.5 km/h (SD = 11.8) less on average (compared to no-train baseline) at around 80 m ahead of LC (A2) In the questionnaire, the respondents were asked to watch the videos and report when they would not anymore start crossing the rails. The minimum safe crossing margin was calculated as the remaining time before the train arrival, determined by one second accuracy as the time when the front of the train reached the right edge of the camera view. The results showed that that crossing margins were shorter in all daytime videos (Mdn = 22 s, M = 28 s, SD = 17 s) compared to the night time videos (Mdn = 84 s, M = 77 s, SD = 30 s). In the night time conditions, the crossing margins were more spread, but also the videos were longer (approximately 60 s vs 120 s). Experts had shorter crossing margins than non-experts (Mdn = 42 s, M = 46 s, SD = 34 s vs. Mdn = 59 s, M = 59 s, SD= 34 s). The main observation is that there were no clear differences between the configurations in the distribution of the time gaps (A3).
Ct20	test-pilot-short	Phase 2 Questionnaire relevant questions: 10. How likely it is that you would ignore the information provided by the in-car alert system (e.g. crossing after being alerted to an approaching train)? Q10: Almost half (46.43%) answered ‘not at all’. Only 7.14% answered ‘considerably’ and no driver chose the option ‘completely’(A4).
Cb23	base-pilot-long	Phase 1 Questionnaire relevant questions: 10. How important is it for you to know how far away the train is from the LC? 11. How important is it for you to know when the train will arrive at the LC?. Those questions received very similar answers. More than 66% answered ‘completely’ and another 15% ‘considerably’ important. Less than 10% stated ‘slightly’ or ‘not at all’ (A4).
Gb19	base-pilot-short	The mean speed of bicyclists on the last 25 m ahead of the LC was *M* = 17.5 km/h (*SD* = 5.58).
Gt20	test-pilot-short	The increased visual search behaviour that was observed with the measure (cf. Gb11 and Gt12) is expected to enhance the information basis for a better decision (not) to cross. The mean speed of bicyclists on the last 25 m ahead of the LC was lower with the measure (*M* = 16.4 km/h, *SD* = 5.54) than in the baseline (A5).
Ib19	base-pilot-short	The mean speed of bicyclists on the last 25 m ahead of the LC was *M* = 17.5 km/h (SD = 5.58).
It20	test-pilot-short	The increased visual search behaviour that is expected for the measure in pedestrians (cf. Ib11 and It12) is also expected to enhance the information basis for a better decision (not) to cross. The mean speed of bicyclists on the last 25 m ahead of the LC with the measure (M = 17.6 km/h, SD = 5.72) was however comparable to that in the baseline (A5).
Jb17	base-literature-short	Stopping behavior (full stop) frequency with no train coming: 4%, with train coming: 41%. Second-step speed reduction (60-2 m in front) with no train coming: 5 mph, with train coming: 10 mph (Liu, Bartnik, Richards, & Khattak, 2016). Only 33% of participants (8 of 24) checked the tracks for a train before crossing (Grippenkoven & Dietsch, 2015).
Jb19	base-pilot-short	Mean speed [km/h] with (SD) on approach to LC at:. 240 m: 53.0 (6.1). - 160 m: 51.9 (6.4). - 80 m: 50.1 (6.2). - 40 m: 47.7 (8.1). - 30 m: 47.0, (9.2). - 20 m: 45.5 (12.3). - 10 m: 46.0 (11.6). - 0 m: 47.2 (9.8). - Allowed max. speed was 50 km/h (A2).
Kb17	base-literature-short	V85 (=speed that cuts off 85% of the observed distribution; i.e. 15% of drivers go faster): 94 km/h at 240 m from LC; 81 km/h at 80 m (allowed max. speed was 100 km/h) (Skládaný, Tučka et al., 2016).
Kt18	test-literature-short	V85 (=speed that cuts off 85% of the observed distribution; i.e. 15% of drivers go faster): 89 km/h at 240 from LC; 69 km/h at 80 m (Skládaný, Tučka et al., 2016).
Kb19	base-pilot-short	Mean speed [km/h] with (SD) on approach to LC at:. 240 m: 53.0 (6.1). - 160 m: 51.9 (6.4). - 80 m: 50.1 (6.2). - 40 m: 47.7 (8.1). - 30 m: 47.0, (9.2). - 20 m: 45.5 (12.3). - 10 m: 46.0 (11.6). - 0 m: 47.2 (9.8). - Allowed max. speed was 50 km/h (A2).
Kt20	test-pilot-short	Mean speed profile on approach to LC: no difference to baseline. Moderate mean value on subjective rating “motivates to drive cautiously”: *M* = 3.6, *SD* = 1.6 (scale: 1-not at all to 6-completely true) (A2).

Ref.	Context	Findings
Kt22	test-literature-long	speed unchanged at LCs with only one group of RSs (also had Give-Way signs); about 8 km/h slower at LCs with four groups of RSs (also had Stop Signs) (Radajj & Kidd, 2005). Speed ~1-2 km/h faster than at short-term after measurement, but still lower than in before phase (Skládaný, Tučka et al., 2016).
Lt18	test-literature-short	Minimal reduction of mean speed on approach to LC in speed study. In observed driving behaviour, 12 of 18 drivers showed “no” reaction (i.e. showed “typical behaviour” with braking at ~100 m from LC and crossing at reduced speed). 6 drivers demonstrated changes in behaviour in “more cautious” direction, no “adverse” reactions observed (Noyce & Fambro, 1998).
Lb19	base-pilot-short	Mean speed [km/h] with (SD) on approach to LC at:. 240 m: 53.0 (6.1). - 160 m: 51.9 (6.4). - 80 m: 50.1 (6.2). - 40 m: 47.7 (8.1). - 30 m: 47.0, (9.2). - 20 m: 45.5 (12.3). - 10 m: 46.0 (11.6). - 0 m: 47.2 (9.8). - Allowed max. speed was 50 km/h (A2).
Lt20	test-pilot-short	Mean speed profile on approach to LC: no difference to baseline. High mean value on subjective rating “motivates to drive cautiously”: *M* = 4.7, *SD* = 1.3 (scale: 1-not at all to 6-completely true) (A2).

3.2.8. Behavioral execution

Table 15. Overview of the HFAT assessment of the safety effects of the piloted measures at the stage of behavioral execution.

	Measure	Score	Evidence				
			Timescale	Literature		Pilot	
				base	test	base	test
A	Blinking lights for locomotive front	2	short-term		At26	Ab27	At28
			long-term				
B	Coloured road markings on approach to LC	NA	short-term				
			long-term				
C	In-vehicle proximity warning (1)	1	short-term				Ct28
			long-term			Cb31	
D	In-vehicle proximity warning (2)	NA	short-term				
			long-term				
E	Rings upstream of the LC	NA	short-term				
			long-term				
F	Traffic light	NA	short-term				
			long-term				
G	Blinking amber light with train symbol	1	short-term			Gb27	Gt28
			long-term				
H	Funnel effect pylons	NA	short-term				
			long-term				
I	Message "Is a train coming?” on road	1	short-term			Ib27	It28
			long-term				
J	Peripheral blinking lights	3	short-term	Jb25	Jt26		Jt28
			long-term				
K	Rumble strips	3	short-term	Kb25	Kt26	Kb27	Kt28
			long-term				
L	Sign Look for train	2	short-term		Lt26	Lb27	Lt28
			long-term				
M	Speed bumps and flashing posts	NA	short-term				
			long-term				

Table 16. Evidence collected on the measures’ effects on behavioral execution.

Ref.	Context	Findings
At26	test-literature-short	Fewer crashes per locomotive/km reported in one study (small sample size) (Cairney, 2003).
Ab27	base-pilot-short	see Ab19
At28	test-pilot-short	see At20

Ref.	Context	Findings
Ct28	test-pilot-short	Phase 2 Questionnaire relevant question: 11.To what extent do you take risks at LCs (e.g. crossing after being alerted to an approaching train)? Q11: Almost all drivers (92.6%) answered 'not at all' and therefore would not take risks around LCs, after being warned by the safety system (A4).
Cb31	base-pilot-long	Phase 1 Questionnaire relevant question: 12. To what extent do you take risks at LCs (e.g. crossing after being alerted of an approaching train)? For Q12, 70.5% answered 'Not at all'. Almost 21% answered 'slightly' and 'moderately'. Only 2.6% stated 'considerably'. However, 7% answered 'completely'. Furthermore, spatiotemporal data about the vehicle kinematics (Floating Car Data) were recorded and analysed to study changes in the behavioural execution of drivers when they approach LCs before and after the measure. The data were processed to form groups of datapoints representing vehicle trajectories through LCs. The data utilized for this analysis were recorded until 15th April 2019, around 2 active and protected level crossings were the active warning pop-up was available. 1846 test vehicles trajectories were identified in the data, 88 of which occurred during the baseline period. 1379 trajectories identified at LC with id=3 and 76 at LC with id=1. The datapoints are map-matched to the street network to calculate the distance to the rail with the minimum possible error. The vehicle trajectories around each LC are aggregated with respect to vehicle's moving direction, since a LC may be approached from two directions. The mean speed and mean acceleration curves as a function of distance to the LC were generated for each LC, direction and period (before and after the measure). The results do not indicate a behavioural change before and after the application of the safety measure. It is found that the typical speed profiles, although differentiated between different LCs and/or approaching direction, are quite similar when comparing a certain LC-direction combination before and after the safety measure. The last 50 meters of vehicle trajectories around LC with id 3 (with the most vehicle trajectories) were further analysed. The GPS pulses were grouped with respect to LC proximity, in 11 levels (0,5, 10,…50 meters distance to the LC respectively). At each proximity level, statistical tests were performed to check correlation between the dichotomous variable period (with values "baseline" and "after") and the continuous variables vehicle speed and acceleration. The appropriate method to examine the association between such types of variables is point biserial correlation, which is valid under several assumptions, two off which were violated in the dataset. Those assumptions are a) no outliers for the continuous variable for each category of the dichotomous variable; and b: the continuous variable should have equal variances for each category of the dichotomous variable. The first assumption, regarding the outliers, was checked by interpreting of boxplots, where outliers in data are outlined. The second assumption was checked by performing the Levene's test, according to which there is a difference between the variances in the two categories. Consequently, there are not sufficient evidence to support the hypothesis that the safety system had a significant change in the behavioural execution of drivers when approaching a LC. This result was, to some extent, expected, considering that the test vehicle drivers are highly experienced, professional taxi drivers who are aware of the locations of LCs and approach LCs with safety (A4).
Gb27	base-pilot-short	see Gb19
Gt28	test-pilot-short	see Gt20
Ib27	base-pilot-short	see Ib19
It28	test-pilot-short	see It20
Jb25	base-literature-short	Average speed on last meter before LC: 38.6 km/h (allowed max. = 50 km/h). Initial speed reductions usually took place around 80 to 50 meters in front of the LC. Drivers who looked for a train decelerated significantly more than those who did not look (33.2 vs. 41.9 km/h on last meter before LC) (Grippenkoven & Dietsch, 2015). Average speed in front of passive LC: Daylight: absolute values not reported – Night: c. 40 km/h (Grippenkoven et al., 2016).
Jt26	test-literature-short	Speed in front of passive LC: Daylight: absolute values not reported – Night: c. 30 km/h (Grippenkoven et al., 2016).
Jt28	test-pilot-short	see Ht20

Ref.	Context	Findings
Kb25	base-literature-short	Different mean speeds observed at the 4 passive LC dependent on signage (Stop vs. give way) and other LC features (e.g. 47.7 to 93.5 km/h at 50 m from LC) (Hore-Lacy, 2008).
Kt26	test-literature-short	Small reductions in mean speed (from -2.7 km/h to -11.6 at 50 m from the LC). No sudden braking observed as potential adverse effect. At sites where rumble strips did not cover the entire road width, some vehicles attempted to avoid them (opposite lane, no oncoming traffic in any of the instances; e.g. 1 of 13 vehicles observed) (Hore-Lacy, 2008).
Kb27	base-pilot-short	see Ib19
Kt28	test-pilot-short	see It 20
Lt26	test-literature-short	see Jb18
Lb27	base-pilot-short	see Jb19
Lt28	test-pilot-short	see Jt20

3.3. Acceptance, trust and usability

The third and last HFAT section (blue) is used to estimate user experience and social perception of the safety measure on the part of road users and other stakeholders. Each measure is assessed in three categories:

- *Acceptance*,
- *perceived reliability*, and
- *perceived usability*.

Within the first category, three further sub categories are distinguished: The *estimated level of acceptance by the public* (e.g. road users, people living near the LC), the *estimated level of acceptance by relevant stakeholders* (e.g. the railway operators, rail infrastructure managers, train drivers, authorities or governments), and the *estimated extent to which the measure can be integrated with the road and rail environment and with other safety measures.* The category *reliability* refers to the estimated extent to which the users of the LC trust the system and perceive it to be fail-safe. *Usability* refers to the estimated extent to which the design of the safety measure is self-explaining (e.g. easy to understand or use) to all road users.

A score is assigned to each measure in each sub category on a scale from *0 (unacceptable)* to *5 (excellent).* For each score, a reasoning is to be provided, indicating the findings or assumptions the score has been based on.

The results obtained in this HFAT section are summarized in Table 17. The scores are colour-coded, with darker shading indicating higher values. In two cases, two scores were assigned for a given measure in a given category. In these cases, the mean of the two values is given in the table. The reasoning behind each score is indicated by a short reference appearing to the right of the score (e.g. *A1)*. The reference consists of a capital letter indicating the measure (e.g. *A -*

blinking lights for locomotive front), followed by a number, indicating the sub category assessed (e.g. 1 - estimated acceptance by public). As reasonings were provided for almost all scores, no extra shading was used to indicate the availability of a reasoning behind the score. All the reasonings referred to in the overview table are listed in a subsequent table.

Table 17. Overview of the HFAT assessment of estimated acceptance, trust and usability.

Measure		Scores and reasoning by sub category									
		Acceptance						Reliability		Usability	
		Acceptance by public		Acceptance by stakeholders		Integration potential		User Trust		Level of self-explaining nature	
A	Blinking lights for locomotive front	3	A1	3	A2	3	A3	4	A4	4	A5
B	Coloured road markings	3,5	B1	2	B2	2	B3	1	B4	1	B5
C	In-vehicle proximity warning (1)	4	C1	5	C2	4	C3	4	C4	4	C5
D	In-vehicle proximity warning (2)	4	D1	4	D2	4	D3	3	D4	3	D5
E	Rings upstream of the LC	4	E1	1	E2	1	E3	2	E4	2	E5
F	Traffic lights	4	F1	2	F2	2	F3	4	F4	4	F5
G	Blinking amber light with train symbol	4	G1	4	G2	4	G3	3	G4	3	G5
H	Funnel effect pylons	0	H1	0	H2	0	H3	0	H4	0	H5
I	Message "Is a train coming?" on road	4	I1	4	I2	4	I3	4	I4	4	I5
J	Peripheral blinking lights	4	J1	4	J2	4	J3	3	J4	4	J5
K	Rumble strips	3	K1	4	K2	4	K3	4	K4	2	K5
L	Sign Look for train	4	L1	4	L2	4	L3	4	L4	4	L5
M	Speed bumps and flashing posts	2,5	M1	3	M2	3	M3	3	M4	3	M5

Table 18. Reasonings provided for the HFAT scores of acceptance, trust and usability.

Ref.	Reasoning behind the score
A1	Concerns about glare reported in one study in literature, possibility of glare at night also expressed by two participants in the pilot study. In the questionnaire, many experts and non-experts expressed concerns on glare. => Design of light device should prevent glare (shades, adaptive intensity, installation height etc.). In the questionnaire, some respondents also indicated that flashing lights may be disturbing and some were concerned about possible misinterpretations of flashing lights.

A2 Reportedly cheaper solution than equipping passive LC with active controls. Design needs to be compliant with existing rail standards
A3 There are some restrictions that need to be considered in design (e.g. certain colours could probably not be used, even if they proved more effective under certain condition [e.g. daylight]; there are regulations about the positions of lights on trains). Still, it is expected that an effective solution can be designed that complies with the restrictions.
A4 The system function might be perceived as technically rather simple and therefore robust and reliable. Three participants in the pilot study expressed the concern that road users might overly rely on the system, which would be a problem in mixed operations with some trains equipped and others not.
A5 System has high face-validity to increase the detection of trains. Strobe lights were judged to be effective and attention-getting in a study (Devoe & Abernethy, 1975; from Carney, 2003).
Two participants in the pilot study noted that the measure would not be effective at LCs with constrained sight. At LCs with highly constrained sight however, passive protection is not the option of choice due to laws and regulations in railway systems (A2).
In the questionnaire, most respondents preferred blinking lights over conventional lights (day time: 94 %, night time: 85 %) (A3).
B1 This measure does not cause any nuisance or constraint to the public. But for many subjects this measure brings no added value compared to the current situation.
B2 For France it depends on each road local authority
B3 For France it depends on each road local authority
B4 Few participants expressed an interest in this measure because it was perceived as duplicating the standard signaling.
B5 The interpretations of the subjects are too heterogeneous.
C1 Acceptance of road users is high, according to their feedback from the 2nd phase questionnaire (Q 14) answers, where over 40% answered that they would be very interested in using the system after the end of the test period and only 10% would not be interested at all. Furthermore (Q 5), 90% of drivers generally feel at least slightly safer using the measure These numbers and results are very promising, taking into consideration that they concern professional (taxi) drivers who are extremely experienced and know the area and LC locations very well. Less experienced drivers might accept the measure to a greater extent.
C2 The measure does not affect the operation of relevant stakeholders. Therefore, they are expected to accept the measure because safety at LC will increase without negative aftermath
C3 The measure is by nature integrated with the road and rail environment, provided that trains are tracked with geolocation devices and road vehicles use navigation software. Those requirements are commonly met, considering the current technology standards.
C4 The system function might be perceived as technically rather simple and therefore robust and reliable. Three participants in the pilot study expressed the concern that road users might overly rely on the system, which would be a problem in mixed operations with some trains equipped and others not.
C5 System has high face-validity to increase the detection of trains. Strobe lights were judged to be effective and attention-getting in a study (Devoe & Abernethy, 1975; from Cariney, 2003). Two participants in the pilot study noted that the measure would not be effective at LCs with constrained sight. At LCs with highly constrained sight however, passive protection is not the option of choice due to laws and regulations in railway systems (A2).
D1 The majority accepts this measure, but some reject it as a source of distraction.
D2 Need to demonstrate the number of collision avoided thanks to this kind of message
D3 Need to demonstrate the number of collision avoided thanks to this kind of message
D4 We think that subjects estimate that it is safe cause they didn't wait to see the LC to begin to reduce their speed
D5 Some messages were not well understood, as *LC in 200 meters*. But other message as *LC closed* or *LC construction works* were well understood.
E1 Subjects appreciate the measure for the design aspect, but not for safety
E2 Need to integrate the measure in regulation to enable deployment
E3 No real impact on road user behaviour
E4 Subjects were confused by this measure and did not link it with the LC
E5 Problem of understanding of the measure and risk of distraction by focus on this measure
F1 This measure is well accepted by road users even though some subjects were confused to encounter the flashing orange light because they were not used to encounter this type of light.
F2 For France it depends on each road local authority
F3 For France it depends on each road local authority

F4 Users better understand the concept of orange and red of a traffic light

F5 Users better understand the concept of orange (danger) and red (stop) of a traffic light

G1 This measure is expected to be well accepted by road users

G2 Expected to be high due to the comparatively low cost and restrictions for installation.

G3 As the system is triggered by approaching road users, not trains, the integration is expected to be uncomplicated. Light should be positioned not to obstruct St. Andrew's cross. Power supply needed.

G4 Expected to be good for users who understand the function. However, if users assume the measure announces a train, trust can become reduced.

G5 No data are available on this, but in general users are expected to make a correct connection between an amber blinking train symbol directly ahead of the tracks and the hazard that a train might be approaching. However, the measure's meaning might be mistaken by some users as actually announcing an approaching train.

H1 Subjects didn’t see the measure

H2 No interest cause no impact on road behaviour

H3 No interest cause no impact on road behaviour

H4 Subjects didn’t link it with the LC

H5 Subjects didn’t understand the measure

I1 No plausible reason to expect that people living nearby could be bothered by the road marking. Road users could be bothered by bad design (e.g. requiring a lot of attentional resources); this can be prevented by considering human factors in road marking design.

I2 Expected to be high due to comparatively low cost and easy implementation

I3 Expected to be high due to low requirements on integration with the railway infrastructure.

I4 Expected to be high due to improbability of technical failure.

I5 Expected to be high, given HF-oriented design.

J1 Finding from Grippenkoven et al., 2016: no differences in lane-keeping, no abrupt reactions (e.g. full braking) in the condition with blinking peripheral lights. Subjective ratings: participants experienced no significant glare or subjective vision impairment from the flashing light device.

J2 Expected to be high due to the comparatively low cost and restrictions for installation.

J3 Given a suitable design and harmonization with the local conditions (e.g. avoiding distraction at critical locations, allowing enough time to react and reduce speed etc.), the prospects for successful integration are expected to be high. Power supply needed.

J4 Due to the autonomous nature of attention capture, user trust does not appear to be a necessary precondition for the measure to work.

J5 Very high because attention capture does not require mental effort and blinking directs attention to the critical regions.

K1 Skládaný, Tučka et al. (2016) report positive attitudes expressed by the interviewees (drivers) towards the measure (“practically not a single bad rating or opinion that the measure was futile”). Still, avoidance behavior was observed in some drivers (avoiding the strips with part of the car width). In a study by Hore-Lacy (2008), no avoidance was observed when RSscovered the entire road width. People living near the LC might be bothered by the enhanced noise level

K2 Given the proof of utility the level of acceptance by stakeholders is expected to be high due to the probable cost advantage of rumble strips compared to classic upgrades. Unclear, if and in what way road maintenance demand might be increased.

K3 Given a suitable design and harmonization with the local conditions (e.g. avoiding distraction at critical locations, allowing enough time to react and reduce speed etc.), the prospects for successful integration are expected to be high.

K4 Expected to be high due to improbability of technical failure.

K5 In Skládaný, Tučka et al. (2016), 34% of respondents expressed the opinion that the LC was not safe for traffic before the intervention. After the installation of rumble strips, this proportion decreased, but still was at 23%. Rumble strips do not by themselves provide directive cues or information and thus can only be explanatory in combination with the surroundings (e.g. signs). In the pilot test, only 31% of participants reported they understood what the rumble strips meant (A2).

L1 No plausible reason to expect that people living nearby could be bothered by the sign. Road users could be bothered by bad design (e.g. requiring a lot of attentional resources); this can be prevented by considering HF in sign design.

L2 Expected to be high due to comparatively low cost and easy implementation

L3	Regulations need to be followed (e.g. in Germany, sign not allowed to be put in direct combination with St. Andrew’s Cross, but could be put in advance), but no severe conflicts to be expected. Sign design and integration with other measures should be optimized for HF aspects.
L4	Expected to be high due to improbability of technical failure.
L5	Expected to be high, given HF-oriented design.
M1	Some subjects stated they "hate" speed bumps because of the bumps representing a hazard to motorcyclists or reducing the comfort of driving
M2	Specific regulation about bumps
M3	The solution cannot be deployed at every LC, its suitability will depend on the road configuration
M4	Subjects do not link it with the LC but know there is a hazard on the road which requires speed reduction
M5	Some subjects stated they "hate" speed bumps because of the bumps representing a hazard to motorcyclists or reducing the comfort of driving

3.4. Summary of results

3.4.1. Applicability of the measures

Concerning the types of LC protection that were selected in the HFAT for each measure, the patterns obtained mostly matched the primary categorization of all, active, and passive LCs, respectively. Two exceptions are observed in the measures *funnel effect pylons* and *speed bumps and flashing posts.* Originally, these were mainly conceived for passive LCs. In the HFAT categorization, their scope was extended to include active LCs with barriers, too, and they were also tested under this use case. This apparent incongruity can be explained when considering the special situation in France where the pilot was carried out. The majority of LCs there is equipped with automatic user side protection and warning (i.e. barriers, 68% of LCs), and most of the LC accidents (72%) happen at this LC type (Silla, Peltola et al., 2017).

With regard to other characteristics of the road and LC environment, the measures assessed can be used under all the circumstances listed, including low and high traffic volume, different qualities of road pavement, as well as sharp and wide crossing angles between road and rails. The only exceptions to this are measures that are applied to paved roads (road markings, rumble strips, and, for the most part, speed bumps) and are therefore not useful in environments with gravel roads. Most of the measures are feasible under a broad range of environmental conditions, including daylight, dusk, dawn, peak traffic hours, rain, as well as urban and rural environments. The few exceptions again involve measures that are applied to paved roads and are hard to discern under conditions of reduced visibility (e.g. darkness, snow). Reduced visibility due to fog affects the efficacy of most of the measures that are applied in the infrastructure as opposed to inside a vehicle.

Only five of the measures address all types of road users: the blinking lights for the locomotive front, the blinking amber light with the train symbol, the peripheral blinking lights, the sign "Is a train coming?" and the road marking with the same message[1]. Almost all of the measures can be applied to address MRUs. Motorcyclists constitute a special group with regard to two types of measures: in-vehicle or mobile proximity warnings and structural changes of the road that are designed to force speed reduction, but also affect motorcyclists' grip on the road and balance (rumble strips, speed bumps). The latter type of measure is not applicable either for VRUs that move at lower speeds, such as pedestrians and bicyclists.

Regarding the adaptation to individual characteristics and conditions of users, all measures can be applied to persons of all genders and ages, except the in-vehicle warning which does not hold for children because they do not take the role of a driver. Concerning reduced sensory abilities, it has to be stated that none of the measures is suitable for road users suffering vision loss, as the majority of measures solely addresses the visual channel. The few ones that use other modalities

[1] Compared to the pilot implementation, the application of the road marking "Is a train coming?" to MRUs requires the adaptation of its design in order to be discernible at faster velocities. This could be done, e.g., by expanding its dimensions and shortening the message (e.g. "← *Train?* →", or a train symbol with the two arrows).

as well (haptic, auditory) do not hold for persons with impaired vision either, as their effect is tied to taking the role of a driver (in-vehicle proximity warning, speed bumps, rumble strips). An additional observation that could inform the further methodological development of the HFAT is that concerning the applicability of measures to persons with special conditions, it seems as if there were two different understandings among the test site leaders filling the HFAT, of when a measure should be indicated as suitable for certain features or states. One understanding was obviously to mark a measure as suitable if it has the potential to facilitate safe behavior in these cases, too. The other seems to have been to mark a measure as suitable only if it specifically addresses persons with the respective individual characteristic or condition. Taking the first of the two perspectives, it should be stated that all of the measures can be effective for persons with impaired hearing and reduced mobility. For persons with intellectual disability, only three measures were attested the potential to be effective. All of these involve the use of blinking lights.

Concerning the intended effect mechanism, up to three mechanisms could be selected per measure by the revisers. The vast majority of measures tested is expected to work by improving the detection of the LC (n = 11) and increasing the user's awareness of the correct behaviour and the consequences of rule violations (n = 12). The next most frequent mechanism is for measures to reduce the approach speeds of vehicles (n = 6). Four of the measures aim at improving the detection of the train, and two measures provide up-to-date information about the status of the LC.

In the next section, dealing with the facilitation of safe road user behavior, the scores that were assigned to the measures will be summarized first, before evaluating the availability of findings to support these scores.

3.4.2. Behavioral safety effects

The results of the assessment in the green HFAT section, behavioral safety effects, are summarized in Table 19. Regarding the earliest stage of information processing, detection and identification, the responsible partners were able to assign a score to each of the measures. This mirrors the fact that for each of the measures, at least one of the intended effect mechnisms is to improve the conspicuity of either the LC or the train. Six measures gained high scores (4 or 5) on this dimension: the blinking lights for the locomotive front, the two kinds of in-vehicle proximity warnings, the traffic light, the peripheral blinking lights, and the speed bumps and flashing posts. Two measures gained low scores (0 or 1): the funnel effect pylons and the message "Is a train coming?” on the road.

A score could be assigned to each of the measures also concerning their effect on the activation of relevant knowledge. This matches the fact that virtually all of the measures are intended to work by increasing road users’ awareness of the correct behaviour and the consequences of rule violation. Four measures gained high scores on this dimension: the blinking lights for the locomotive front, the in-vehicle proximity warning no. 2, the peripheral blinking lights, and the sign „Is a train coming“. Two measures gained low scores: the funnel effect pylons and the in-vehicle proximity warning no. 1.

Table 19. Overview of the scores and evidence on behavioral safety effects.
(*Time* – timescale, *s* – short-term, *l* – long-term, *X* – evidence available)

Measure	Time	Detection & Identification					Rule Knowledge					Decision-Making					Behavioral Execution				
		Score	Lit.		Pilot		Score	Lit.		Pilot		Score	Lit.		Pilot		Score	Lit.		Pilot	
			base	test	base	test		base	test	base	test		base	test	base	test		base	test	base	test
Blinking lights for locomotive front	s	5	X	X	X	X	4	X	X		X	4	X	X	X	X	2		X	X	X
	l																				
Coloured road markings on approach to LC	s	3				X	3				X	NA					NA				
	l																				
In-vehicle proximity warning (1)	s	5				X	1				X	4				X	1				X
	l				X					X					X					X	
In-vehicle proximity warning (2)	s	5				X	4				X	NA					NA				
	l																				
Rings upstream of the LC	s	3				X	2				X	NA					NA				
	l																				
Traffic light	s	4				X	3				X	NA					NA				
	l																				
Blinking amber light with train symbol	s	3				X	3			X	X	2			X	X	1			X	X
	l																				
Funnel effect pylons	s	0				X	0				X	NA					NA				
	l																				
Message "Is a train coming?" on road	s	1				X	2			X	X	1			X	X	1			X	X
	l																				
Peripheral blinking lights	s	4	X	X	X	X	4	X	X		X	4	X		X	X	3	X	X		X
	l																				
Rumble strips	s	2	X	X	X	X	2	X	X		X	2	X	X	X	X	3	X	X	X	X
	l		X	X										X							
Sign Look for train	s	3		X	X	X	4		X		X	4		X	X	X	2		X	X	X
	l																				
Speed bumps and flashing posts	s	4				X	3				X	NA					NA				
	l																				

Regarding the stage of decision-making, only seven of the measures were assigned a score. This may seem surprising, considering the high occurence of measures intended to increase the user's awareness of the correct behaviour and the consequences of rule violation. However, as this intended effect mechanism category combines aspects relevant to both knowledge activation (*user's awareness of the correct behaviour …*) and decision-making (*… consequences of rule violation*), one reason for this may be that the piloted measures focus on the first aspect more than on the latter. Of the measures that received a score concerning the facilitation of decision-making, four scored high on this dimension: the blinking lights for the locomotive front, the in-vehicle proximity warning no. 2, the peripheral blinking lights, and the sign „Is a train coming“. A low score was yielded by one measure: the message "Is a train coming?” on the road.

Concerning behavioral execution, again, only seven of the measure were assigned a score. No measure scored high on this dimension. Three measures gained a low score: the in-vehicle proximity warning no. 1, the blinking amber light with train symbol, and the message "Is a train coming?” on the road. The remaining four measures gained medium scores. None of the measures was assessed as having adverse effects on safe behavior on any stage of information processing.

Looking at the evidence on which the scoring was based, each score assigned (not *NA*) has at least one reference to an observation or reasoning it is based on: for every measure, there is a reference to results from the pilot, observed in the test condition on a short timescale. For seven measures, moreover, a description of baseline behaviour is reported from the pilot, from a short- or long-term perspective. Evidence from earlier studies could be gathered for four measures only.

3.4.3. Acceptance, trust and usability

Concerning the *estimated level of acceptance by the public* (e.g. road users, people living near the LC), the majority of measures gained either high scores (4 or 5; $n = 8$) or at least medium scores (2 or 3; $n = 4$). Only one measure, the *funnel effect pylons*, was assigned a low score (0), supported by the reasoning that it would not be accepted by road users because they did not notice it. Overall, the values assigned on this dimension are slightly higher than on the other acceptance and trust dimensions.

Regarding the *estimated level of acceptance by relevant stakeholders* (e.g. the railway operators, rail infrastructure managers, train drivers, authorities or governments), again, the majority of measures scored either high ($n = 7$) or within a medium range ($n = 4$). Two measures were expected to be hardly accepted: for the *funnel effect pylons*, this was due to missing evidence of their impact on road behaviour; for the *rings upstream of the LC* due to the need to integrate the measure in the existing regulation to enable a deployment. The results for the *estimated extent to which the measure can be integrated with the road and rail environment and with other safety measures* basically mirror the results in the dimension before. This suggests that revisers might have taken the ease or difficulty of integration as a major source of expected stakeholder acceptance.

User trust as the *extent to which the users of the LC trust the system and perceive it to be fail-safe* was also mostly expected to be high ($n = 6$) or at least moderate ($n = 5$). Two measures gained low scores on estimated user trust. One of them was the *funnel effect pylons,* for reasons already mentioned. The other one was the *coloured road markings*. The reasoning behind this score was that few participants felt interested in this measure because it duplicated the signaling. As the measure was tested at an active LC in the pilot, it remains unclear whether this observation can be generalized to other application contexts.

Usability, as the *estimated extent to which the design of the safety measure is self-explaining (e.g. easy to understand or use) to all road users*, was mostly expected to be high ($n = 6$) or moderate ($n = 5$). Low scores were assigned again to the *coloured road markings* and the *funnel effect pylons*, as participants in the pilot did not reliably understand the meaning of these measures.

3.5. Overall assessment

3.5.1. Integration of the results

Now, which measures score best from a human-factors perspective? Before trying to answer this question, we would like to introduce some basic considerations on the integration of the HFAT results. So far, the HFAT does not specify a quantitative procedure on how all the information obtained should be considered and integrated in an overall assessment – and there are some good reasons for this. In the development of the HFAT, it was discussed whether it was reasonable to determine a sum score from the single scores obtained on the measures' effects on information processing. The answer was no. Pertinent features of human information processing and its relation to behavioral outcomes make a qualitative consideration and integration of the information obtained appear more reasonable, at least at this stage of development.

It is hard to find a universally valid quantification of how each of the stages of human information processing is associated with the outcome of safe behavior. This is because the importance of each of the stages is not always the same, but varies with the requirements that a defined task imposes on a given individual in a defined context. To describe the tasks and contexts and also some of the individual features that are relevant to assess a certain LC safety measure, the qualitative information obtained in the first HFAT section can be consulted.

One major factor influencing the task demands imposed on a road user at a given LC is the type of protection applied there. As an example, passive LCs, unlike active ones, demand of road users to determine by themselves whether it is safe to cross. Therefore, special emphasis lies on the first stage of detection and identification that needs to be facilitated by a safety measure. However, the outcome at this stage is not independent of other stages. For example, the activation of rule knowledge ("This is a passive LC – this means I need to check by myself whether a train is coming.") is one way by which processes of detection and identification can be improved (Corbetta & Shulman, 2002; Yantis, 2000). An alternative way would be to get the user to look to the left and right by applying the exogenous capture of attention that is autonomously elicited by flickering stimuli in the periphery of the visual field (ibd.; Itti & Koch, 2000). In this case, the road user does not need to activate relevant knowledge about LCs, but the effect is mediated by the behavioral execution of eye movements. Thus, the same effect – improvement of detection – can be achieved by facilitation on two different stages of information processing. The role of the different processes is further diversified by other user and context features, as for example road user type. This variable is associated with a number of features that modify task demands, such as a certain viewpoint, a typical velocity range, and certain other sensory and motion capacities (e.g. diving through below closed full-barriers for pedestrians vs. cars). Taking the typical velocity as another example, detection at passive LCs could also be facilitated by measures that induce road users to slow down (behavioral execution) and thus give them more time to process information and react on LC approach – but this only holds for road users who typically move at high speeds. Moreover, although slowing down improves the conditions to detect an approaching train and come to a stop, it is neither a necessary, nor a sufficient condition for that road users indeed check to the left and right for a train in time.

These examples show that it is not reasonable to require a certain LC safety measure to score high on the facilitation of *all* stages of information processing in order to be assessed as effective. Moreover, there are connections between the stages of information processing, similar results may be achieved through different paths, and the processes at some stages may not be as important for certain road users or contexts as for others. A simple sum score would not satisfactorily reflect this situation. One potential solution of devising an overall score could be the attempt to specify requirement profiles for a selection of defined prototypical use cases (combinations of task, user and context features) that weight the relevance of certain dimensions in a given use case. However, the combinations that would need to be mapped in such a numerical weighting system, are numerous. Considering the HFAT as a way of collecting and systemizing relevant information, a comparative assessment can also be attained in a qualitative approach. Such an approach will be proposed and used to integrate the HFAT results in the next section.

3.5.2. Comparative evaluation of the measures from a human-factors perspective

The applicability information obtained in the first HFAT section is mainly to provide context information to allow better assessment of the results in the other two sections: It provides the background information to judge the behavioral safety effects and the prospects for acceptance and trust. However, the first section can also contribute by itself one information relevant to an overall assessment, which concerns the versatility of a measure. From a human-factors perspective, great versatility is not needed concerning all those variables that can be specified in advance in the selection of a given measure for a given LC site – e.g. LC type, LC characteristics, urban vs. rural setting, and, if applicable, the definition of certain road user types encountered there. In theses cases, versatility is not necessary because a safety measure can and should be chosen according to its designated operation conditions. Great versatility is however desirable concerning all factors that are still subject to change after a measure has been chosen for a given LC, which mostly concerns the *feasibility at different times of the day, under different weather conditions, for different types within the relevant set of road users encountered there, different individual characteristics and states.* This versatility could be used as one criterion in a comparative assessment of different measures.

The piloted measures do not differ much in terms of the defined kind of versatility. Based on the considerations in section 3.1, two groups could be distinguished: One group with slightly smaller versatility, including measures that are applied to paved roads (road markings, rumble strips, and speed bumps), and another group including the other measures.

Concerning the assessment of behavioral safety effects, the first step proposed in a comparative evaluation should be to look at the availability of evidence to underpin the scores assigned. An estimation of this can be captured at a glance from Table 19, by focusing on the “X”s in the shaded cells. A quality criterion here is for a measure to have evaluation results available from a lot of different contexts. The more information is available, the more substantiated the numerical scores can be considered. Among the piloted measures, the *rumble strips* are the best-evaluated measure: Besides the results from the pilot test, there is also evidence available from earlier studies, including long-term evidence from test conditions. Next, the *blinking lights for the locomotive front,* the *peripheral blinking lights,* and the *sign “← Is a train coming? →”* are comparatively well-evaluated, with evidence also including insights from other studies, however without findings related to long-term effects. Finally, there is a group of measures for which the only evidence available at the moment originates from the pilot tests. Within these, for the *in-vehicle proximity warning no. 1*, the *blinking amber light with a train symbol*, and the *message "← Is a train coming? →" on the road,* there are data available from a baseline condition without a measure applied, to which the effects of the measure can be compared. No baseline is described concerning the effects of the *coloured road markings on LC approach*, the *in-vehicle proximity warning no. 2*, the *rings upstream of the LC*, the *traffic light*, the *funnel effect pylons*, and the *speed bumps and flashing posts*. As these were tested in the same pilot, it is however possible to draw a comparison between the outcomes obtained with each of the measures.

Having a very well-evaluated measure does not by itself imply that the measure is "good", from a human-factors perspective. Instead, it means that the scores assigned to this measure are well-substantiated and therefore relatively reliable and little dependent on subjective evaluations. Likewise, scores with less evidence may hit the mark, but there is more uncertainty associated to them. As an analogy facilitating the handling of the results, the amount of evidence available could be thought of in terms of error margins around the score that are narrower for scores with a lot of evidence and wider for scores with relatively little evidence.

Next, the importance of each of the stages of information processing in terms of a safe behavioral outcome should be reflected for each measure. This can be done by taking into account information from the applicability section of the HFAT, to define what processes the measure is supposed to enhance for a given road user at a given LC. In the first phase of SAFER-LC Task 2.3, a *psychological function involved* was defined for the effect of each measure. However, as this classification was mainly thought for the purpose of sorting the measures, only one or maximum two main functions were defined per measure. Moreover, for a number of measures, the two functions were *detection* and *identification*, which are now included in one stage. For the purpose of considering the importance of the stages in an overall comparison of the measures, this does not appear sufficient, minding the interactions between the stages. Therefore, the information on the main psychological function involved was complemented with the information on the *intended effect mechanisms,* of which three had been indicated per measures in the applicability section of the HFAT (

Table 7). Based on this information, the stages of information processing that are most relevant to each measure were highlighted in Table 19 in bold font.

All of the measures are intended to improve the detection of the LC and / or the train; therefore the stage of *detection and identification* is an important touchstone for each measure. For all measures that are intended to increase the user's awareness of correct behaviour and consequences of rule violation, both *rule knowledge* and *decision-making* were marked, as aspects of both are included in this effect mechanism. For measures providing up-to-date information about LC status, the *rule knowledge* stage was marked, as the up-to-date information mainly affects the knowledge available to direct behavior. For measures intended to reduce the approach speeds of vehicles, the stage of *behavioral execution* was highlighted. In the subsequent comparison of the measures in terms of their behavioral safety effects, the focus is put on the scores at the relevant stages of information processing only for each measure.

Finally, acceptance, trust and usability should be considered in an overall assessment, too, because they create the framework conditions for the measures to unfold their effects on road user behavior. Especially low levels of acceptance, trust and usability can act as barriers to the efficacy of measures. To take account of this in the integration of results, the last step after the comparison of the behavioral safety effects and the consideration of evidence is to look out for low values on the acceptance and trust dimensions and the reasons behind them.

According to the scores on the relevant stages, the four measures assessed to most facilitate safe road user behavior were the *blinking lights for the locomotive front*, the two *in-vehicle proximity warnings,* and the *peripheral blinking lights*. Minding the evidence available, this assessment is rather certain for the two measures involving blinking lights (with the restriction that the longevity of effects has not yet been proved), and more tentative for the *in-vehicle proximity warnings.* Acceptance and trust are expected to be sufficient to allow for successful implementation of these measures, minding the principles of user-friendly design and stakeholder participation.

Two measures scored particularly low on the assessment of behavioral safety effects: the *funnel effect pylons* and the *message "← Is a train coming? →" on the road*. Both assessments are tentative, as the findings from the pilot are the only evidence available by now. Due to the low expected efficacy, acceptance and trust values are not considered in these cases.

The seven remaining measures were attested medium effectivity for the facilitation of safe behavior, according to their intended ways of working. These medium scores are more certain for the *rumble strips* and the *sign "← Is a train coming? →",* and remain tentative for the *coloured road markings on approach to LC*, *the rings upstream of the LC*, the *traffic light*, the *blinking amber light with a train symbol*, and the *speed bumps and flashing posts*. Acceptance and trust are expected to be sufficient to allow for successful implementation for most of these measures, except the *coloured road markings on approach to LC*, the *rings upstream of the LC*, and the *funnel effect pylons*. It should be noted here that in the pilot involving the coloured road markings and the funnel effect pylons, it is likely that the rendering of these two measures in the simulator did not reach a satifsfactory fidelity level. This should be born in mind in the interpretation of all results concerning these measures.

4. DISCUSSION

4.1. Methodological considerations

4.1.1. Strengths and limitations of the assessment method

Based on pilot tests that involved the assessment of data on road user behavior and experience, thirteen human-centered LC safety measures were evaluated from a human-factors perspective. Of these, three measures were conceived for use at all LC types, three mainly for use at active LCs with barriers, and seven mainly for use at passive LCs. The human factors assessment involved the description and evaluation of each measure with regard to its applicability, its effects on road user behavior at four different stages of information processing, and its expected acceptance and social perception by road users and other stakeholders. A designated tool (HFAT) was used for the assessment.

The HFAT is a way of collecting and systemizing relevant information on the effects of LC safety measures in terms of human factors. Moreover, it is a useful guide in the definition of study designs to assess these effects. Its structure and requirements convey essential methodological principles, such as the consideration of existing research evidence, the comparison of data from a baseline and a test conditon, the use of a control condition in before-after designs, the definition of relevant behavioral dimensions to be assessed, and the recognition of the importance of acceptance and trust for efficacy. By using the evidence collected as an input to subjective assessments on a scale from 0 to 5, the HFAT allows to integrate the results of very different studies. As any research method, the use of the HFAT has some limitations that shall be addressed at this point.

The quality and certainty with which the quantitative assessment scores can be assigned depends both on the quantity and quality of inputs from the preceding step of collecting evidence from empirical studies. Therefore, the tool requires its user to carefully collect empirical evidence. To allow the assessment of the available evidence and / or the derivation of prognoses in case of a lack of evidence, the HFAT also requires basic knowledge of human behavior and knowledge in the field of road and rail transport.

Using expert assessment as a way of integrating data from different sources in a numerical score, the HFAT relies on the revisers' consideration of the available evidence. Although experts bring in a lot of knowledge and experience, the scores are always subjective assessments. On the one hand, this is a strength of the method because it allows to achieve a score even based on fragmentary evidence or in the absence of empirical findings, based on theoretical considerations. On the other hand, the scores are necessarily affected by features of the revisers, e.g. their previous knowledge on research related to similar safety measures, the effectiveness of safety measures, level crossing safety etc. This is all the more true, the less evidence is available.

In the work presented here, the HFAT was used by each of the different test site leaders to document and evaluate the results of their own pilot. This was a pragmatic and parsimonious approach to yield an HFAT assessment of each measure. At the same time, it makes the comparison of the results across measures more difficult, because of potential individual differences in the interpretation of the HFAT requirements and the meaning of the research findings attained. Examples of this can be found in apparent differences in the comprehension of how to use the applicability category on individual characteristics and conditions of users, or the handling of missing data.

4.1.2. Feedback for the further development of the HFAT

A first revision of the HFAT has already been done between the first and the second assessment phase (cf. section 2.3). Another revision is planned to complete the research activities in the SAFER-LC project by using the feedback from the second assessment phase and the evaluation in the further optimization of the human factors methodological framework, including the HFAT. The following lessons learned and suggestions can be derived from the current evaluation of human-centered low-cost measures.

In its current form, the tool already contains detailed instructions of how it should be applied, including an examplary completion of the forms for one measure. However, further details were discovered in the evaluation in terms of how these instructions could be refined. With regard to the first HFAT section, applicability, two apparently different interpretations were observed among the revisers concerning the sub categories *disability, under influence of,* and *under skill impairing states* within the indicator subset *adaptation to individual characteristics and conditions of users*. One apparent understanding was that a measure should be indicated as suitable for a certain user characteristic if it has the potential to facilitate safe behavior in persons with this characteristic, too. The other one was probably to indicate a measure as suitable only if it specifically addresses persons with the respective characteristic. A revised HFAT could further specify how this categorization should be handled.

Further specifications could also address the second HFAT section, behavioral safety effects. An explicit standard could be defined of how to deal with the assignment of a score to stages to which no finding can be directly allocated: Should a score be assigned based on theoretical reasoning, or should no score be assigned at all? As a related topic, an explicit standard could be defined of how to deal with stages that are not directly influenced by a measure. One possible way would be to reason that an influence on one stage can also affect the processing at following stages (e.g. if a user is more likely to detect a train approaching, this may also influence her decision not to cross the tracks in a critical phase). Another way would be to require that a 0 should be assigned to the following stages if they were not measured by independent indicators. Whatever method is chosen should fit the method of integrating all the results in the end.

The first version of the HFAT made a distinction in the early stages of information processing between *detection* and *identification*, analogous to the model of human information processing (Grippenkoven and Dietsch, 2015; Havârneanu et al., 2018). In the review phase after the first assessment it was decided to integrate the two stages into one, due to the observation that measures that help detection mostly also help identification, and behavioral indicators as they can be collected in applied research mostly make it hard to clearly assign the results to just one of the two stages. Moreover, the theoretical distinction can be difficult to handle to revisers that are no human factors experts. The experience of using the HFAT on the results of the pilot studies confirms the feasibility of integrating these two stages into one assessment.

Considering the results in terms of the assignment of findings to the stages of information processing, another way of facilitating the use of the HFAT could be to include more specific behavioral descriptions of the target effects on behavior within the stages. For example, evidence concerning an observed speed reduction was sometimes cited for *decision-making*, sometimes for *behavioral execution*. If there was a behavioral indicator as, e.g., “induces speed reduction on approach”, it would be easier for the revisers to find the right place to insert a finding from a study. Moreover, a more specific description could be helpful in the specification of requirement profiles, too (see below). For example, knowing that reduced speed on approach of a passive LC is a good prerequisite to enable effective visual search and coming to a stop in time if necessary, but futile if it is not indeed combined with increased visual scanning, this could be reflected in the requirements.

In the current evaluation, a qualitative approach was used to integrate information from all the three HFAT sections in an overall assessment of measures. This approach could be used as a starting point to further refine the integration of results in a future procedure. As mentioned already, it may be possible to devise a reasonable procedure for the computation of an overall score if the relevance of single stages (or indicators, see above) can be defined a priori for a given measure, based on its scope (application context), and these relevance values can be used as weights in the computation of the overall score. In the current analysis, the scope of the measure was considered by using the information on the main psychological functions and intended effect mechanism. This may be refined by including further information such as the target LC type (e.g. Is approach speed relevant to accidents there?) and the target road users (e.g.: How fast are they usually?) in the definition of relevant stages or indicators (e.g.: Is speed reduction a desirable target behavior?). In order to avoid an abundance of possible combinations that would need to be parametrized in this way, the analysis could start with a few prototypical use cases in order to assess whether is a promising approach. These use cases could be selected as combinations of road user type, LC type and potential other features that are specifically relevant in LC accidents (cf. Silla et al., 2017)

Finally, the reliability of the scores could be assessed in a test-retest and / or inter-rater comparison. The first approach would involve presenting the same revisers again with the evidence collected and asking them to assign scores; the second to present other revisers with the evidence and letting them do the scoring. The values obtained could then be compared to the original scores to assess the precision of the assessment tool.

4.2. Practical implications

The human factors assessment reported here focused on the suitability of measures in their defined application context. That is, the research question was how well a given measure can support safe road user behavior at exactly the kind of LC it is conceived and designed for. Economic aspects were reflected only in terms of the requirement for measures to be low-cost, with low-cost measures defined as measures costing less than a classic upgrade when applied to a large number of LCs (cf. section 1.3). Furthermore, the assessment did explicitly not address the measures’ overall potential contribution to road and rail safety in terms of accidents prevented, which considers the absolute frequency of different LC types in addition to the efficacy of a measure to promote safe behavior at a given LC type. Estimates of this can be found in Silla et al. (2019).

The selection of piloted measures covers the most important types of LC protection and is applicable in a wide range of road and other environmental conditions. The measures are clearly human-centered, which is also expressed in the intended effect mechanims. They are innovative in the sense that they are not already in common use to protect LCs in the European countries. In terms of the low-cost requirement, there are some differences between piloted measures with regard to the cost dimension. A detailed analysis of costs and benefits is carried out in WP5 of the SAFER-LC project.

Most of the piloted measures address motorized road users. Five of the measures can be effective for all user types, including vulnerable road users. The feasibility of measures for individual characteristics and conditions of users is generally good, with the exception that none of the measures is suitable in its piloted form for road users suffering vision loss. The needs of road users with reduced sensory abilities need to be reflected in measure design. For the piloted measures, this adaptation could be done by including information in the auditory and haptic modality (e.g. acoustic infrastructure signals, tactile paving), or using mobile communication to convey personalized information.

The measures assessed to most facilitate safe road user behavior were blinking lights for the locomotive front, in-vehicle proximity warnings, and peripheral blinking lights at the LC. The two former ones can be applied at all kinds of LCs, the latter is for use at passive LCs. The scores for the two measures involving blinking lights are supported by the results of multiple studies including the pilot tests, while the score for the in-vehicle proximity warnings is more tentative with the only evidence available by now coming from the pilot test.

A shortcoming observed in the availability of evidence on the effects of the piloted measures is that long-term evidence is generally scarce. For the SAFER-LC pilots, this resulted from the constrained time frame and limited resources in data collection and analysis. The same difficulty is observed in a lot of other scientific studies. Long-term assessment marks a transition of measures from development to practice as it requires a durable integration of new measures into existing infrastructures. Therefore, possibly long-term evidence must be drawn from trials of measures that show good prospects on the timescales tested, by communities and rail infrastructure managers.

On a theoretical basis, for in-vehicle proximity warnings, some habituation effects can be expected in the long term, because, to be effective, the measure requires a voluntary effort of the driver to initiate the recommended behavior. The autonomous capture of visual attention by flickering stimuli in the periphery of the visual field, as used in the blinking train and the peripheral blinking lights, is a hard-wired feature of the nervous system. This automatism evolved because it represented an evolutionary advantage. Therefore, this reaction is not expected to be subject to considerable habituation effects.

In the research reported here, the HFAT was used as a tool to guide the planning of empirical studies on the effects of LC safety measures and to allow the structured documentation of the results obtained. However, the tool has further application potentials. To road and railway stakeholders, it can serve as a checklist to support the consideration of human factors aspects in the evaluation of LC safety measures. By going through the HFAT forms and dealing with the questions and assessments with respect to a given measure, all the major human factors aspects of importance will be considered. This includes the different potential impacts on road user behavior as well as expected acceptance and social perception. Using the HFAT in this checklist function can help to assess the suitability of a LC safety measure to different railway environments and user groups, and to create the necessary conditions for efficacy by considering the perspectives of road users and other stakeholders.

The results obtained in SAFER-LC Task 2.3, the design and evaluation of human-centered low-cost measures to improve LC safety, will be used as one main input in the implementation of the *SAFER-LC toolbox*, a web-based tool for anyone concerned with LC safety, as road and rail infrastructure managers, train operators, engineers, designers, scientists, decision-makers, policy makers and standardisation bodies rail and road managers. The toolbox is conceived to be a guide to best practice that integrates all the recommendations, promising interventions, and specifications developed during the project with the empirical evidence collected from the scientific literature and the pilot tests. The toolbox will be accessible free of charge at the end of the project and will continue to be maintained, updated and improved by the International Union of Railways (UIC) for the benefit of the entire road- and railway-safety community.

5. REFERENCES

- Cairney, P. T. (2003). Prospects for improving the conspicuity of trains at passive railway grade crossings. Washington, D.C: Transportation Research Board.
- Carroll, A. A., Multer, J., & Markos, S. H. (1995). Safety of Highway-Railroad Grade Crossings: Use of Auxiliary External Alerting Devices to Improve Locomotive Conspicuity. Report DOT-VNTSC-FRA-95-10. Washington, DC.
- Corbetta, M. & Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience. 3 (3): 201–215.
- DB Netze. (2018). Bahnübergänge im Spiegel der Statistik - Bahnübergangsstatistik 2017: Version für Dritte. Berlin, München: DB Netz AG / TÜV Süd Industrie Service GmbH.
- Dreßler, A., Silla, A., Kortsari, A., Havârneanu, G., Whalley, S., Lorenzo, L., & Grippenkoven, J. (2018). Definition of new human-centered low-cost countermeasures: Deliverable D2.3 of the SAFER-LC project. Available from http://safer-lc.eu/IMG/pdf/saferlc_20181031_d23_v10web_dlr_definition_of_new_human-centred_low-cost_countermeasures.pdf
- European Union Agency for Railways (2019). ERAIL – European Railway Accident Information Links. Available from https://erail.era.europa.eu/
- Grippenkoven, J. (2017). Wahrnehmung und Verhalten am Bahnübergang. Deine Bahn, 2, 10–15.
- Grippenkoven, J., & Dietsch, S. (2015). Gaze direction and driving behavior of drivers at level crossings. Journal of Transportation Safety & Security, 8(sup1), 4–18.

- Grippenkoven, J., Thomas, B., & Lemmer, K. (2016). PeriLight – effektive Blicklenkung am Bahnübergang. EI - Der Eisenbahningenieur, (1), 48–51.
- Havarneanu, G., Silla, A., Whalley, S., Kortsari, A., Dreßler, A., & Grippenkoven, J. (2018). Human Factor methodological framework and application guide for testing (interim report): Deliverable 2.2 of the SAFER-LC project. Available from http://safer-lc.eu/IMG/pdf/saferlc_20180724_d22_v04_uic_hf_methodological_framework.pdf
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research, 40.
- Knake-Langhorst, Sascha und Gimm, Kay und Frankiewicz, Tobias und Köster, Frank. 2016. Test Site AIM – Toolbox and Enabler for Applied Research and Development in Traffic and Mobility. Transportation Research Procedia, 14, Seiten 2197-2206. Elsevier. DOI: 10.1016/j.trpro.2016.05.235.
- Liu, J., Bartnik, B., Richards, S. H., & Khattak, A. J. (2016). Driver behavior at highway–rail grade crossings with passive traffic controls: A driving simulator study. Journal of Transportation Safety & Security, 8(sup1), 37–55.
- Silla, A., Lehtonen, E., Virtanen, A., Boufidis, N.; Salanova Grau, J.; Dressler, A.; Grippenkoven, J.; Taillandier, V.; Khoudour, L.; Jacqueline, D.; Antoine, R.; Boukour, F.; Edelmayer, A.; Ruffin, C.; Kassa, E. & Korkmaz, C. (2019). Results of the evaluation of the pilot tests. Deliverable D4.4 of the SAFER-LC project
- Silla, A., Peltola, H., Aittoniemi, E., Sintonen, H., Kortsari, A., Taillandier, V., ... (2017). Level crossing accidents and factors behind them: Deliverable 1.2 of the SAFER-LC project. Available from http://safer-lc.eu/IMG/pdf/saferlc_20171003_d12_v04_vtt_lc_accidents_and_factors_behind_them.pdf
- Silla, A., Seise, A., & Kallberg, V-P. (2015). *Survey and assessment of measures aiming to improve the safety of level crossings*. Finnish Transport Agency, Infrastructure and Environment. Helsinki 2015. Research Reports of the Finnish Transport Agency 7/2015. [Finnish with English summary]. Available from http://www2.liikennevirasto.fi/julkaisut/pdf8/lts_2015-07_tasoristeysten_turvallisuustoimenpiteiden_web.pdf
- UNECE. (2016). Assessment of safety at level crossings in UNECE member countries and other selected countries and strategic framework for improving safety at level crossings. Retrieved from https://www.unece.org/fileadmin/DAM/trans/doc/2016/sc2/ECE-TRANS-SC2-2016-id05e.pdf (17-04-2019)
- Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. In Monsell, S.; Driver, J. (eds.). Control of cognitive processes. Attention and performance. 18. Cambridge, MA: MIT Press. pp. 73–103. ISBN 978-0-262-13367-8. OCLC 44039778.

ANNEX A1: METHODS OF PILOT DRIVING SIMULATOR STUDY (SNCF)

All measures were tested in the driving simulator of SNCF. The main aim was to identify the effects of the implemented safety measures on driver behaviour by comparing the driving behaviour before and after implementing these measures. The evaluation was conducted based on within subjects before–after measurements. The comparison was made between a crossing behaviour of a conventional LC and a LC with a measure.

The baseline data was collected during 3–4 minutes drive in the city centre in a route without any LCs but including a STOP sign, traffic lights, give way situation, a roundabout and a road outside agglomeration with different speed in a straight line and in a curved road. In practice, this was a route beginning of the actual test drive in the simulator.

The actual test drives were 20–30 minutes long and the route included both open and closed LCs (seven situations with LC closed and six situations with LC open). The first three LCs in the route were 'classic LCs' which were considered as control condition. After that, the participants began to meet LCs with equipped with a different safety measures.

Information regarding the following variables was collected for the evaluation (sensors to collect these information were activated at all situations):

- Speed sensor of the vehicle compared to speed limit
- Videos of driver in the vehicle and video of simulation (to compare the time lapse)

In addition, after test drive, each simulator participant was interviewed during 30 to 50 minutes by a cognitive expert according to Vermesch's method. This qualitative explicitation interview allows the subjects to verbalise the mental or physical actions implemented in a situation. The main interview topics were:

- Perception
- Reasoning
- Action
- Comprehension

In total, 58 persons participated in the simulator study. Of these, 33 persons participated in the safety course and 25 persons participated in the connected vehicle course. Out of all 58 participants, five were professional drivers.

The test subjects were selected based on various criteria for gender, age, occupation, number of years holding a driving license as well as typical trip purpose (work/home or professional appointments) and its frequency:

- 53% of subjects were women and 47% men
- 14 subjects were 14−24 years old and 19 were 25−35 years old (25 subjects were 35−50 years old (of which at least half of them had children)
- Five were professional drivers (commercial, taxis, technicians, etc.)

- 10 subjects reported their annually driving less than 5000 km; 20 subjects 5000−20000 km and 10 subjects more than 20000 km
- In terms of the number of years holding a driving license, seven subjects reported less than two years, eight subjects between two and five years and 43 subjects reported between five and 30 years

For the coloured road markings and the funnel effect pylons, it is likely that the rendering of the simulator did not adequately represent the quality hoped for in reality. Therefore, the results obtained are to be considered with caution, and the evaluation of this safety measure should be done in real road environment.

ANNEX A2: METHODS OF PILOT DRIVING SIMULATOR STUDY (DLR)

All four measures were tested in the MoSAIC driving simulator of DLR[2].

Table A1. *Procedure and time plan of the driving simulation study.*

Trial / Phase	Duration	Contents	Factor 1 (within subj.): *LC measure*	Factor 2 (within subj.) *Train presence*	Factor 3 (between subj.) *Train design* *(nested in „train coming")*		
A	5 min	*Welcome and instruction*					
B	5 min	*Informed consent*					
C	2 min	*explanations in simulator*					
D	8 min	*calibration of eyetracking system*					
0	5 min	*training drive*	**no LC**	**no train**			
1	7 min	***Baseline*** *test (always first)*	**no measure** (=control / baseline)	**no train** (=control / baseline)			
2	7 min	***Effects of Factor 1*** *- Position of measure balanced across subjects*	**Blinking PeriLight <-> Noise-prod. pavement <-> Sign** *Look for train*	**no train** (=control / baseline)			
3	7 min						
4	7 min						
5	7 min	***Effects of Factor 2*** *- only one train design per subject*	none	**train coming**	**Normal** (=baseline for train-specific comparisons)	or	**Blinking Lights on train** to enhance train detection
6	7 min	***Effect of train exposition*** *- additional LC traverse for testing the effect*	**no measure** (= experimental condition after train exposition)	**no train**			
E	18 min	*Survey of subjective data on the scenarios experienced (5 or 6)*					
F	3 min	*Debriefing*					
G	2 min	*Disbursement and farewell*					
Driving time	47 min			**target n subjects**	18		18
Total duration	**90 min**			**target total n**	**36**		

[2] https://www.dlr.de/fs/en/desktopdefault.aspx/tabid-11368/19984_read-46631/

A lot of effort was spent on the design of a driving environment the allows studying multiple measures in a within subjects experimental design, meaning that the same participant can be confronted with multiple measures without making them suspicious that the study is about level crossing safety. A long driving course was planned, consisting of both village sections and rural roads between them. To distract the participant from the level crossing focus of the study, a secondary task had to be completed once while driving through each of the villages in between the LCs. For this, participants received a message on a mobile phone, requiring them to execute a small task and send a short reply to the enquirer (e.g. “Please find the photo of the electric kettle I wanted to put on ebay and send it to me”). The secondary task was part of the cover story used to justify the purpose of the study in the initial instruction. Participants were briefed on the real purpose after the study.

A detailed procedure of the studies and the stepwise sequence of events is shown in Table A1. After a phase of introduction, explanation and calibration (A–D), participants started with a training course to get used to the simulator. The subsequent test course contained six LCs and was designed to take about 7 min driving time from one LC to the next. The first LC-passage always served as a baseline. A passive LC was crossed without a train approaching.

The second to fourth LC likewise entailed a passage without a train coming, but with one of three different infrastructural safety-measures in place (*sign ‘Look for train’*, *peripheral blinking lights* or *noise-producing pavement*). These three experimental conditions were encountered by all of the participants. The order of measures was balanced across participants.

At the fifth level crossing, each participant encountered a train that approached the LC. This train was a normal train for a half of the participants (baseline for the factor *train design*), and a train with additional blinking lights for the other half. While the other factors were varied within-subjects, a between-subjects design was chosen for this factor because a train encounter was expected to bias driver behaviour in any following LC passages towards more attentiveness and caution than would normally be observed at LCs. Therefore, each participant should only encounter a train once, and always at the end of the measure sequence. The direction of train approach from the left vs. right side was balanced across participants. The train was triggered when the participant’s car was at 390 m ahead of the LC, with its trajectory to be on a perfect collision course in case the driver would continue to drive at the maximum allowed speed of 50 km/h.

To test for the effects of a train encounter on driver behaviour at LCs, a sixth LC passage was included, involving a LC without any supplemental safety measure and without a train, similar to the baseline condition.

After the test-drive, a questionnaire was administered to the participants in which they were first briefed on the background and focus of the study and then shown each of the measures again, along a with a short description of their proposed functions. Participants were subsequently asked to give their assessment of the measure on a number of scales.

A total of 52 participants (24 male, 28 female) took part in the study. The conduct of the study and the assessment of the driving, gaze and subjective data, respectively, was partially restricted due to simulator sickness (participants were instructed to abort the test immediately in this case), technical problems with gaze detection or calibration quality in eye-tracking, and, in one case, persisting failure to comply with the instructions. Participants who had to quit early because of simulator sickness, still filled in the user questionnaire if they felt ok to do so. Subjective assessments were collected of 49 participants (24 male, 25 female, aged 18 to 65, M = 35.3, SD = 13.1). A complete set of driving data could be obtained of 46 participants (22 male, 24 female, aged 18 to 65, *M* = 34.4, *SD* = 12.5), and a complete set of gaze data was obtained of 39 participants (18 male, 21 female, aged 18 to 65, *M = 34.4*, *SD = 12.7*).

To assess the effect of the measures on visual search for a train, indicators of looking out for a train on the tracks to the left and right were computed and compared between the conditions. The necessary stopping distance at a speed of 50 km/h (including reaction and braking) is about 40 m with normal braking, or 30 m with hazard braking. Therefore, the analysis focused on gaze behaviour in the LC approach section from 140 to 40 m ahead of the LC, in which visual scanning for a train is especially important to determine whether there is a need to brake and give way to a train.

Figure 3 shows the regions of interest (ROI) that were defined as the left periphery and right periphery for the analysis. Fixations needed to last at least 120 ms to be counted. For the defined ROI and approach phase, the following indicators were computed: (1) number and percent of participants who fixated the ROI at least once, and (2) mean number of fixations on the ROI.

Figure 3. Definition of the regions of interest (ROI) "left periphery" (L) and "right periphery" (R). The blue dot represents the participant's current gaze position in this screenshot.

In the condition involving a train encounter, the train was triggered on the left or right side when the participant's car passed a fixed trigger point ahead of the LC. It first became visible on the screen on average when the participant's was at 250 m ahead of the LC. The trigger point was chosen to achieve a situation in which the participant's car would have to give way to the train when approaching at a speed of 50 km/h. Therefore, the train was already present on the screens and hence detectable before the participant's car reached the critical approach region that was analysed in the ROI analyses reported above. To test whether the blinking train was detected earlier than the normal train the *time of first fixation* on the train in terms of distance ahead of the LC was analysed for each participant.

The analysis of drivers' speed choices on LC approach focuses on the 300 m ahead of each LC, i.e. it starts at the point at which the first sign of the respective LC infrastructure (three-striped post at 240 m ahead of the LC) becomes discernable, and ends at the beginning of the LC (0 m). To assess the effects of the different measures, we look at the velocity difference between each condition and the baseline – i.e., how much slower or faster did drivers go on average at a certain point with a certain measure compared to the situation without the measure –, and compare the resulting difference profiles across the measures.

ANNEX A3: METHODS OF PILOT VIDEO-BASED SURVEY (VTT)

The additional warning light system was tested at real railway environment both from the viewpoint of road user and engine driver.

The tests were conducted on 14th March in Sääksjärvi in Finland. The testing was done in main railway network and one of the three tracks was reserved for the tests. No official level crossing existed at the test site. However, it was a location where the road user camera could be easily installed (two meters from the track around 1.25 meter height).

The rented railway vehicle was driven through the imaginary level crossings several times both in day time conditions and during darkness. The approach of the railway vehicle to the imaginary level crossing was video recorded both from the angle of the road user (from the road side) and from the angle of the train driver.

The variables included in the tests are presented in Table A2. The speed of the railway vehicles during the tests were 20 km/h. In addition, the possible annoyance of additional warning lights were estimated both from the road user and engine driver perspective.

Table A2. *Variables investigated during the tests.*

Title	Variable
Time of day	– Daylight (12:00–13:30) – Night (at 11 pm–1:30 am)
Light configuration at daytime. Two runs for each scenario.	– Reference with standard lights – 1 x 100 ms flash in every 2 second – 2 x 100 ms flash in every 2 second – 3 x 100 ms flash in every 2 second – 1 + 2 + 3 100 ms flash in every 2 seconds
Light configuration at night time. One run for each scenario.	– Reference with standard lights – 1 x 100 ms flash in every 2 second – 2 x 100 ms flash in every 2 second – 3 x 100 ms flash in every 2 second – 1 + 2 + 3 100 ms flash in every 2 seconds – Dimmed lights 2 x 100 ms flash in every 2 seconds – 5° tilt upwards 2 x 100 ms flash in every 2 seconds – 10° tilt upwards 2 x 100 ms flash in every 2 seconds
Perspective	– Road user – Engine driver

The evaluation was carried out with a web-based questionnaire by rail and road safety experts connected to the SAFER-LC project. For comparison, the questionnaire was filled by non-experts.

Three alternative light configurations were compared to the standardly used reference configuration, both in the day time and in the night time conditions. The reference configuration had standard train headlights: three continuous white lights, two on the bottom and one on the top. In the alternative configurations, additional blinking LED lights were installed below each of the headlights. The alternative configurations had different blinking patterns (Table A3).

Table A3. *Four configurations tested.*

Configuration/Number of blinks	Description
0	Reference system without strobe lights
1	Single blink every 1 s
2	Double blink every 2 s
3	Triple blink every 3 s

The questionnaire focused on the expert evaluation of the alternative configurations regarding safety, comfort and suitability for day and night time conditions, as well as on the ergonomical aspects visibility and glare. Benefits and drawbacks were also explicitly asked, and which configuration the experts preferred. Additionally, we investigated if the blinking lights would make the approaching train to appear faster or threatening, and thus influence the judgement of the last safe crossing times.

The questionnaire was based on the road user view videos filmed from the test site. In total, eight videos were used. Four in the day time conditions demonstrating the reference system and the three alternative configurations, and similarly four in the night time conditions. The duration of videos was 66–68 s for the day time videos, and 111–130 s for the night time videos. The night time videos were longer because we wanted the train to become visible in the beginning of the video, and in the night time this occurred earlier.

First all four day time videos were presented and evaluated, followed by the four night time videos. The reference configuration without blinks (0) was always presented first. It was followed by configuration with two blinks (2), configuration with one blink (1), and finally configuration with three blinks (3).

With the reference configuration, the participants were asked to watch the video and report when they would not anymore start crossing the rails. The minimum safe crossing margin was calculated as the remaining time before the train arrival, determined by one second accuracy as the time when the front of the train reached the right edge of the camera view.

For all the alternative configurations, the participants reported similarly the crossing margin, but they were also asked if they saw any benefits or drawbacks with the alternative configuration compared to the reference configuration, and if they did, describe those. For each alternative configuration they were also asked to rate the alternative configuration on safety, comfort, suitability for day/night time conditions, visibility, and glare, using a 5 step Likert scale, where 1 = worse than the reference system, 3 = equivalent to the reference system, and 5 = better than the reference system. After going through all the four day/night time videos, the participants also reported which one they preferred and why.

Finally, the questionnaire asked participants' background information such as age, gender and self-rated expertise on level crossing and road safety, as well as views in improving level crossing safety in general.

Answering to the questionnaire was voluntary and anonymous. The experts' questionnaire was sent via project email list whereas the non-experts' questionnaire was sent to various email lists of the local university. In total, 18 expert and 16 non-expert responses were received and analysed.

ANNEX A4: METHODS OF PILOT LIVING LAB (CERTH, TRAINOSE & DLR)

A before–after study design was used to assess the safety effects of this measure. The before data consisted of 1.5 months of baseline data (situation before the application was in use). During this period, the application was installed and logged spatiotemporal data for the floating taxis near LCs included in the pilot, without producing alerts ('inactive' mode). The data collected in inactive mode were used for assessing the behaviour of drivers around LCs without the safety warnings.

The length of after data collection period was eight months[3]. During this period, the service was fully operable ('active' mode), and data analysis for this period focused on two differentiated cases: static alerts for the proximity of the LC and dynamic alerts for the proximity of a train, issued when a train is expected to reach a LC within a minute.

More than 600 taxis (out of approximately 1,000 taxis operating for the same taxi association) used the application in the city of Thessaloniki, Greece. Taxi drivers were allowed to withdraw from the pilot and have all the data recorded for the vehicles erased, by uninstalling the application at any time. According to the taxi association, some taxis use rather basic tablets with low-end specifications (e.g. 1GB RAM) which struggle to cope with the existing dispatching software and they were expected to not install/use our application, which was a limitation for the testing of the service and it has reduced the performance to lower levels. However, at the same time these have become more representative of a large-scale implementation, in which various users may not have high-processing devices.

The drivers that participated in the program were provided with a written instruction form during the process of application installation. Its purpose was to highlight that application users should never fully entrust the system about the dangers and proximity of trains and that they are fully responsible for taking all necessary safety precautions when driving close to level crossings. On a technical level, the geolocation tracking, data transmission and popup alerts operate autonomously after the mobile application is installed, therefore no further training was required for the application users.

In total, 29 level crossings and various trains in the line Athens–Thessaloniki were included in the pilot. The trains were equipped with GNSS devices monitored by the Greek national train operator TRAINOSE and CERTH-HIT was granted real time access to the train location and speed data.

Besides the safety impact assessment by means of Floating Car Data (FCD), the piloted measure was evaluated in terms of operational performance and user's experience, utilising operational data automatically recorded by the system and questionnaires answered by test vehicle drivers before and after their experience with the piloted measure, respectively. In addition, three taxis

[3] In deliverable D4.3 (Carrese et al., 2019) the data collection period was planned until the end of July, but it was decided to extend this period until mid-September in order to collect evaluation data spanning over a longer period.

were equipped with Naturalistic Driving Study (NDS) equipment to collect data for analysing the drivers' reaction to the safety service in the context of the approach to level crossings. The NDS platform consisted of a set of four miniature cameras. It monitored the environment as well as the driver's behavior and facial expressions during November and December 2018. In addition to the cameras, a GPS sensor was implemented in the NDS system to detect driving parameters such as speed, acceleration and position of the taxis. Four different drivers drove these NDS equipped taxis.

In summary, the datasets recorded and utilised in the evaluation of this measure are the following:

- Vehicle location and speed data generated by trains and taxis
- Data recorded by the safety system backend server
- Questionnaires answered by the drivers of the test vehicles (taxis)
- NDS data

ANNEX A5: METHODS OF PILOT FIELD TEST AT REAL-WORLD LC (DLR)

Test site

The pilot study took place at a passive LC situated in the north of Braunschweig (see Figure A7) mainly frequented by cyclists and pedestrians. The road is closed to four-wheelers, but can also be used by single-track motorized vehicles such as motorbikes, as well as other VRUs like horse riders, wheel-chair users, skaters etc. Leading through a surrounding of meadows and forest, it is used by numerous cyclists on their way to and from work and is also a popular route for leisure trips.

a b

Figure A7. *The test site: passive LC at Ottenroder Straße, Braunschweig (a - aerial view, b - western approach view).*

Equipment

To examine the effects on road user behavior, the DLR mobile traffic data acquisition system was installed at the LC. The system is part of the DLR test field AIM (*Application Platform for Intelligent Mobility*, Knake-Langhorst et al., 2016). The implementation used in the pilot consisted of a semi-mobile pole on a concrete foundation, a sensor head, and a weather-proof cabinet, containing processing computers as well as devices to allow remote access by an LTE-connection and V2X-ability (see Figure A8). The sensors used were a set of stereo-cameras, supported by an active infrared lighting system for artificial scene illumination to enable sensing during day and night time. The system fuses the sensor data and automatically processes them into trajectories of the moving traffic objects detected. The data contain information about the dimensions and classification (e.g. train, pedestrian, cyclist) of the object as well as its location, velocity and other dynamic state variables. The trajectories were tracked with a rate of 25Hz and automatically stored in a database. Moreover, the low-resolution scene videos that are the input to the computation of the trajectories were recorded in accordance with data protection regulations to allow the study of road user behavior beyond kinematics.

a b

Figure A8. *The mobile traffic data acquisition system as used in the pilot: a – pole with sensor head and control cabinet, b – positioning of the system relative to the LC (viewed from rear).*

The traffic data acquisition system was also used to control the elicitation of the blinking in the amber light measure dependent on the approach of VRUs to the LC. For this, a geofencing algorithm was applied to the trajectory data in real time (see Figure A9). The target road segment started at 40 m ahead of the LC and ended at 6 m ahead of it. When a road user was detected entering from the eastern side (right side in the figure), the blinking was triggered, continuing until the road user left the target area. If other road users entered while the amber light was still active, it remained active until the last one was out of the area. The blinking was only elicited by road users with west-bound trajectories; road users traveling in the opposite direction did not influence the amber light.

Figure A9. *Geofencing for triggering the blinking of the amber light (see text for procedure).*

Design and procedure

The pilot data were collected from mid August to the end of September 2019. The pilot started with the assessment of baseline data at the LC with no additional measure applied (2 weeks), followed by a test period with the pavement message '*Is a train coming?*' (2 weeks). After a recovery phase without a test measure (1 week), the blinking amber light was implemented (2 weeks).

Trajectory data

Overall, the trajectory data of 18,529 VRUs on a west-bound trajectory were recorded during the baseline phase and the two test phases. The majority of these VRUs were bicyclists (n = 16,049). The number of pedestrians observed was 2,480. Table A4 shows the frequencies of the types of VRUs split by the three pilot phases.

Table A4. Frequencies of VRU types observed during the test phases.

Condition	VRU type	
	Bicyclists	**Pedestrians**
Baseline	**4,598**	**618**
Message	**6,362**	**861**
Amber light	**5,089**	**1,001**

For slow-moving VRUs such as pedestrians, velocity choices were not expected to play a major role for their possibilities to come to a stop ahead of the LC in due time if necessary. Therefore, velocity ahead of the LC was only analysed as a safety indicator for the VRU group of bicyclists. Using daytime data (between 7:00 am to 7:00 pm), the mean velocity on the last 25 m ahead of the LC was computed for each bicyclist.

Video Annotation Data

To assess VRU behavior on approach to the LC, a sample of the low-resolution videos was annotated using the ELAN software (Version 4.7.3, https://tla.mpi.nl/tools/tla-tools/elan/, Wittenburg et al., 2006). Defined categories of target behaviour included *lateral head movements to the left* and *right* as an indicator of gaze direction, and *visual distraction* (e.g. VRU looks down, looks to other people). VRU features coded included gender, age group, and VRU type. For each of the three test conditions, the behaviour of 80 VRUs with west-bound trajectories was coded. Within each condition, an equal sample was taken from the first and last weekday of the respective phase, and within each of the sampled days, an equal sample was taken from each of the peak times starting at 7:30 a.m. and 5:00 p.m. The first 20 VRUs appearing in each of the defined timeslots were coded. The defined *LC approach* zone started at around 20 m ahead of the tracks and ended at around 1 m ahead of the tracks. The low-resolution videos did not allow observation of VRU behaviour ahead of this zone. Moreover, due to vegetation, VRUs were probably unable to see the track periphery much before entering this area.

The resulting sample consisted of 240 VRUs (n = 133 male, n = 106 female, n = 1 not assigned). Of these, 157 were identified as adults (18–65 years), 37 as youngters (14–17 years), 34 as children (0–13 years), and 12 as seniors (> 65 years). The most frequent road user type observed was bicyclist (n = 214), the remaining VRUs were pedestrians (n = 20), motorcyclists (n = 2), horse riders (n = 1), and other VRUs (n = 3).

To analyse the effects of the applied measures on active visual search for a train, we assessed how many of the observed VRUs turned their head in a given direction (*left*, *right*, *both ways*) at least once on LC approach, or turned their head *neither way*. Head turns that could be assigned to a distraction (e.g. VRU looked to other VRU) were not counted.

ANNEX A6: THE HUMAN-FACTORS ASSESSMENT TOOL

Fill in the following forms for a given safety measure under evaluation. Each form is colour coded to reflect the three different sets of criteria under assessment: the 'Classification criteria' are included in a classification checklist (orange form). The 'Criteria to assess the behavioural safety effects' are included in five separate assessment sheets, one for each criterion (green forms). The 'Criteria to assess the user experience and social perception' are included in one assessment sheet (blue form). Detailed instructions are provided in the forms' headers.

Name of the measure being assessed	Name of the pilot test and brief description of the tested measure
…………………………………	…………………………………………………………………………………………

CLASSIFICATION CRITERIA			
Factor	**Brief description**	**Indicator** ***(Tick only the cases that the measure applies to, or click the option 'All' if the measure covers all the cases)***	
Applicability to different LCs	*Specify the types and characteristics of LCs where the measure can be implemented (multiple answers are possible)*	Type of LCs	☐ All ☐ Passive LCs without any warning device ☐ Active (manual) ☐ Active LCs with half barriers ☐ Active LCs with full barriers ☐ Active LCs with skirts for pedestrians ☐ Active LCs with light and sound warning ☐ Active LCs with other warning device ☐ Active LCs with traffic lights
		Characteristics of LCs	☐ All ☐ LCs with low vehicle traffic ☐ LCs with high vehicle traffic ☐ LCs with paved road ☐ LCs with gravel road ☐ LCs with availability of electricity ☐ LCs with low usage / not used at all ☐ LCs with sharp / wide crossing angle ☐ Other (specify)……………………………
Feasibility under different environmental conditions	*Specify the environmental circumstances in which the measure aims to be most effective and which may affect the perception or the behavioural adaptation of road*	Time of the day	☐ All ☐ Daylight ☐ Darkness ☐ Dusk ☐ Dawn ☐ Peak traffic hours
		Weather conditions	☐ All ☐ Rain ☐ Snowfall ☐ Slipperiness

	users (multiple answers are possible)	Setting of the LC	☐ Fog ☐ Bright sunshine/ glare ☐ All ☐ urban ☐ rural
Applicability to different types of user	*Specify the categories of LC users who are targeted by the measure (multiple answers are possible)*	MRU VRU	☐ All ☐ cars ☐ motorbikes / mopeds ☐ trucks / heavy vehicles ☐ buses / coaches ☐ farm / agricultural vehicles ☐ other (specify)………………………… ☐ All ☐ pedestrians ☐ cyclists ☐ other (specify)…………………………
Adaptation to individual characteristics and conditions of users	*Specify if the measure is applicable for people with the following characteristics or conditions (multiple answers are possible)*	Gender Age Disability Under influence of Under skill impairing states	☐ Male ☐ Female ☐ All ages ☐ children ☐ elderly ☐ vision loss and blindness ☐ hearing loss and deafness ☐ intellectual disability ☐ reduced mobility ☐ other (specify)………………………… ☐ alcohol ☐ drugs ☐ medication ☐ fatigue ☐ stress ☐ Risk-seeking personality
Intended effect mechanism	*Specify the mechanism via which the measure is expected to have an effect on safety (maximum 3 options can be ticked; <u>undeline the main</u> effect mechanism)*	☐ Improves the detection of train ☐ Improves the detection of LC ☐ Controls access to and supports egress from LC ☐ Reduces the approach speeds of vehicles ☐ Increases the user's awareness of correct behaviour and consequences of rule violation ☐ Improves the physical environment of LC ☐ Improves the possibilities of vulnerable road users to cross LC safely ☐ Provides up-to-date information about the status of LC ☐ Supports the LC safety actions ☐ Makes waiting time more tolerable ☐ Other (specify) ……………………………………	

CRITERIA TO ASSESS THE BEHAVIOURAL SAFETY EFFECTS OF MEASURES ON ROAD USERS (SHORT- AND LONG-TERM)

Criterion	Brief description
Detection and Identification	The measure can help the LC user detect relevant visual and auditory stimuli and identify relevant information in the environment which can increase their detection of the LC, an approaching train or other potential danger

Write down brief descriptions of the road user's detection and identification of relevant LC safety information (e.g. detection of LC or train) before and after the measure (including any numerical findings from pilot tests or literature to support the estimated behavioural changes)

Period	Evidence from literature		Evidence from pilot test	
	Short-term*	Long-term*	Short-term*	Long-term*
Before / Without the measure				
After / With the measure				

* Refer to the Application Guide for examples of what can be considered a short- and long-term change

Answer the following question by choosing one score between 0 and 5 or the answer 'N'. Make the choice based on the descriptions you gathered above.

Question: To what extent does the measure facilitate the detection and identification of the LC, train or possible danger while the user is approaching the LC?

Answer modalities	N	The LC user's visual or auditory perception can be impeded/distracted by this measure
	0	This measure has no intended influence on the visual or auditory perception of the LC user
	1	
	2	
	3	
	4	
	5	LC users can easily detect the LC or the approaching train with sufficient time to stop or to cross safely (and continue to do so in the long term)
Score	…	*Reasoning behind the score / Assumption on the short and long-term change in road user behaviour* ……………………………………………………………………………………………

Criterion	Brief description
Rule knowledge	The measure can help the LC user elicit and retrieve relevant information about the required safe behaviour to cross the LC

Write down brief descriptions of the road user's ability to elicit and retrieve relevant safety information before and after the measure (including any numerical findings from pilot tests or literature to support the estimated behavioural changes)				
Period	Evidence from literature		Evidence from pilot test	
	Short-term*	Long-term*	Short-term*	Long-term*
Before / Without the measure				
After / With the measure				

* Refer to the Application Guide for examples of what can be considered a short- and long-term change

Answer the following question by choosing one score between 0 and 5 or the answer 'N'. Make the choice based on the descriptions you gathered above. **Question: To what extent does the measure evoke the required behaviour while the user is approaching the LC?**		
Answer modalities	N	The LC user is confused about how to behave safely at LC, because the measure transmits unclear or misleading information
	0	This measure has no intention to remind the LC user the required/safe behaviour
	1	
	2	
	3	
	4	
	5	LC users understand how to cross the LC safely without prior knowledge or experience of the LC type and environment in question (in all situations, also in the long term)
Score	…	*Reasoning behind the score / Assumption on the short and long-term change in road user behaviour* …………………………………………………………………………………………

Criterion	Brief description
Decision-making	The measure can help the LC user take more accurate decisions that arrive at safe behavioural intentions

Write down brief descriptions of the road user's decisions before and after the measure (including any numerical findings from pilot tests or literature to support the estimated behavioural changes)				
Period	Evidence from literature		Evidence from pilot test	
	Short-term*	Long-term*	Short-term*	Long-term*
Before / Without the measure				
After / With the measure				

* Refer to the Application Guide for examples of what can be considered a short- and long-term change

Answer the following question by choosing one score between 0 and 5 or the answer 'N'. Make the choice based on the descriptions you gathered above. **Question: To what extent does the measure facilitate the user's decision-making towards a safe course of action while approaching the LC?**		
Answer modalities	N	The LC user decides to cross unsafely, because this measure encourages their inaccurate subjective judgment of risk
	0	This measure has no intended influence on the subjective decision-making factors of the LC user
	1	
	2	
	3	
	4	
	5	LC users decide to cross the LC safety, because they understand the risks and the associated consequences of their behaviour (in all situations, also in the long term)
Score	…	*Reasoning behind the score / Assumption on the short and long-term change in road user behaviour* …………………………………………………………………………………………

Criterion	Brief description
Behavioural execution	The measure can 'force' the LC user execute safe actions (required behaviours) or can impede the LC user from executing risky actions (non-adapted behaviours)

Write down brief descriptions of the road user's behavioural execution before and after the measure (including any numerical findings from pilot tests or literature to support the estimated behavioural changes)				
Period	Evidence from literature		Evidence from pilot test	
	Short-term*	Long-term*	Short-term*	Long-term*
Before / Without the measure				
After / With the measure				

* Refer to the Application Guide for examples of what can be considered a short- and long-term change

Answer the following question by choosing one score between 0 and 5 or the answer 'N'. Make the choice based on the descriptions you gathered above. **Question: To what extent does the measure directly influence the safe execution of the approach and crossing behaviour?**		
Answer modalities	N	The ability of the LC user to cross safely is made difficult by this measure
	0	This measure has no intended direct influence on the LC user's execution of actions
	1	
	2	
	3	
	4	
	5	LC users are physically impeded from illegally crossing the LC or are forced to cross the LC safety when this measure is in place (also in the long term)
Score	…	*Reasoning behind the score / Assumption on the short and long-term change in road user behaviour* ……………………………………………………………………………………………

CRITERIA TO ASSESS THE USER EXPERIENCE AND SOCIAL PERCEPTION

Choose the most appropriate answer by ticking one box for each case							
Factor	Definition	(0) Un-acceptable	(1)	(2)	(3)	(4)	(5) Excellent
Accep-tance	The estimated level of acceptance by the public (e.g. road users, people living near the LC)	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
		Reasoning behind the score (indicate the findings or assumptions the score has been based on):					
	The estimated level of acceptance by relevant stakeholders (e.g. the railway operator, rail infrastructure manager, train drivers, authorities or Government)	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
		Reasoning behind the score (indicate the findings or assumptions the score has been based on):					
	The estimated extent to which the measure can be integrated with the road and rail environment and with other safety measures	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
		Reasoning behind the score (indicate the findings or assumptions the score has been based on):					
Reliability	The estimated extent to which the users of the LC trust the system and know that it is fail-safe	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
		Reasoning behind the score (indicate the findings or assumptions the score has been based on):					
Usability	The estimated level of self-explaining nature of the design of safety measure (e.g. easy to understand or use) by all road users, all age categories and persons with various disabilities	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
		Reasoning behind the score (indicate the findings or assumptions the score has been based on):					