IJR International Journal of Railway
Vol. 2, No. 2 / June 2009, pp. 50-59

The Korean Society for Railway

Automated Assessment System for Train Simulators

Marcus Schmitz* and Christian Maag**

Abstract

Numerous train operating companies provide training by means of driving simulators. A detailed analysis in the course of the rail research project 2TRAIN has shown that the simulation technology, the purposes of training and the overall concept of simulator-based training are rather diverse (Schmitz & Maag, 2008). A joint factor however are weak assessment capabilities and the fact that the simulator training is often not embedded into the overall competence management. This fact hinders an optimal use of the simulators. Therefore, 2TRAIN aims at the development of enhanced training and assessment tools. Taking into account that several simulators are already in use, the focus lays on the extension of existing simulation technology instead of developing entirely new systems. This extension comprises (1) a common data simulation interface (CDSI), (2) a rule-based expert system (ExSys), (3) a virtual instructor (VI), and (4) an assessment database (AssDB). The foundation of this technical development is an assessment concept (PERMA concept) that is based on performance markers. The first part of the paper presents this assessment concept and a process model for the two major steps of driver performance assessment, i.e. (1) the specification of exercise and assessment and (2) the assessment algorithm and execution of the assessment. The second part describes the rationale and the functionalities of the simulator add-on tools. Finally, recommendations for further technical improvement and appropriate usage are given based on the results of a pilot study.

Keywords : *Train simulator, Virtual instructor, Rule-based expert system, Railway training, Performance markers.*

1. Introduction

The establishment of safe, competitive, and interoperable railways in Europe requires particular attention to the qualification of train drivers (Elms, 2001; European Commission, 2007; Maag, Schmitz & Fröschl, 2009). In order to strengthen a European harmonisation in rail traffic, it is appropriate to advance common training concepts as well as common training technology. Developments concerning these two aspects are the scientific and technological objectives of the EU-funded rail research project 2TRAIN. The individual European countries differ in national laws, engine technology, signalling systems, rule books, and general training structures. As a consequence of this diversity, a complete harmonisation of training technology and training contents will be unachievable. Nevertheless, as Europe grows together and cross-border operations increase there is a strong need to harmonise and coordinate the education and training of train drivers concerning general driving and operational abilities as well as particular crisis management competencies. Taking into account experiences made in the past, 2TRAIN aims at developing solutions for an efficient, safety-enhancing, and cost-effective use of modern technologies for training as well as for the ongoing competence and performance assessment.

Important benefits can be obtained by the use of computer-based systems and simulators in training (Rail Safety & Standards Board, 2007). These computer-based technologies facilitate the establishment of common training efforts for train drivers and enable an enhancement of training efficiency by conducting interactive training of realistic situations (European Commission, 2007). In contrast to training in real environment, the training session can variably be composed of different technical failures and hazardous operational situations of which many cannot be trained in reality. Furthermore, it is possible to replicate scenarios at any given time. Several safety-related and economic advantages of simulation contribute to the wide distribution of this training method. In the near future, railway

* University of Wuerzburg, Centre for Traffic Sciences (IZVW)
E-mail: schmitz@psychologie.uni-wuerzburg.de
**Roentgenring 11, 97070 Wuerzburg, Germany
E-mail: maag@psycholgie.unii-wuerzburg.de

companies will even intensify the usage of simulators for the training and assessment of train drivers (Schmitz, Maag & Mera, 2008). Besides their usage for training, driving simulators can also be used for railway related research. Especially the study of rail human factors is a research application that is suited for simulation as a research instrument (Dvorak & Brunner, 1987; Groeger, Bradshaw, Everatt, Merat & Field, 2001; Oed & Müller, 1995).

Starting point of 2TRAIN was a benchmarking of training technology, training contents, and training models already in use in Europe. All in all, the analysis highlights the important role of simulation for the training and assessment of train drivers. Nevertheless, the interviewed companies state some shortcomings that hinder an optimal use of the existing training technology. Major points of criticism are the missing possibility to create new exercises and scenarios as well as the weak assessment capabilities of the simulators, in particular the suboptimal quality of the simulator reports. There is also a wide variation amongst European railway undertakings with regard to how they build up, monitor and assess competence of staff.

An optimal benefit from simulator training can be achieved, if it is embedded into the overall competence management of the company. In addition, the simulator exercises should be tailored to the intended learning objectives. It is important that the train drivers become familiar with the new technology by using the training tools regularly in the course of initial training, advanced training, and competence check. In addition, an ongoing performance assessment is necessary to measure the current status of competence and to derive implications for further qualification needs. To achieve a standardised assessment procedure, it is absolutely essential to revert on objective assessment parameters as distinctly measureable performance markers. These objective parameters have to be defined as direct measures of countable behaviour patterns or results (Bommer, Jonson, Rich, Podsakoff & Mackenzie, 1995). Therefore, it is necessary to develop an assessment concept in order to guarantee a reliable comparison of the target behaviour that derives from the rulebooks and the actual behaviour during the simulator training. The 2TRAIN concept is based on performance markers (PERMA concept) combining the usage of simulator data as a precondition for an objective and detailed assessment of the train drivers' performance and the observations by the instructor. Standardised scenarios and procedures for all training centres of a company also lead to a transparent training and assessment.

Based on this assessment concept, simulator add-on tools have been developed to reach a harmonisation of today's diverse training technologies and to achieve a standardised driver assessment (Schmitz & Endres, 2008). In the course of 2TRAIN the extended simulation system is demonstrated at three simulators in Germany (ICE), France (TGV) and Spain (Tram). The scenarios in the three pilots contain similar events as well as specific ones depending on the given system. The target behaviour and the related rules are country and system specific.

2. Assessment and data concept based on performance markers (PERMA concept)

If driving simulators are used for research and training applications, standardised scenarios and procedures are needed in order to get as objective and valid data as possible. The behaviour levels that could be measured by the simulator are (1) actions carried out by the driver and the order of actions, (2) reaction times, and (3) drivers' performance (compliance according to operational rules and regulations).

2.1 Assessment based on performance markers

In the field of simulation-based training and research there are two different assessment approaches: (1) Subjective assessment by the instructor/trainer (i.e. observation) and (2) objective assessment by using simulator data (Hall & Brannick, 2009). Both assessment approaches have advantages and disadvantages–also depending on the concrete training or research objectives. Important benefits of subjective assessment are that no specific technical tools for data collection are needed and an overall assessment of the driver's performance is possible. In contrast to an automated solution it is widely accepted by the train driver, because it is more common for employees to be judged by another person than by a technical system. Many human factor trainings can only be assessed by subjective ratings of the real instructor, e.g. communication, decision making, or dealing with stress. On the other hand, subjective assessment has some disadvantages that are mainly related to a lack in standardisation and reliability. Research has shown that the objective assessment of train drivers by using simulator data is more precise than the subjective assessment by an instructor (Maag, Schmitz, Siebers & Krüger, 2005) even if the instructors are well-trained and supervised. An objective performance assessment could significantly decrease the amount of errors made by the observer during the assessment procedure–especially if the observational task is very broad or requires a high temporal resolution. Independent from the chosen assessment approach the performance assessment should always be based on observable behaviour patterns, i.e. performance markers.

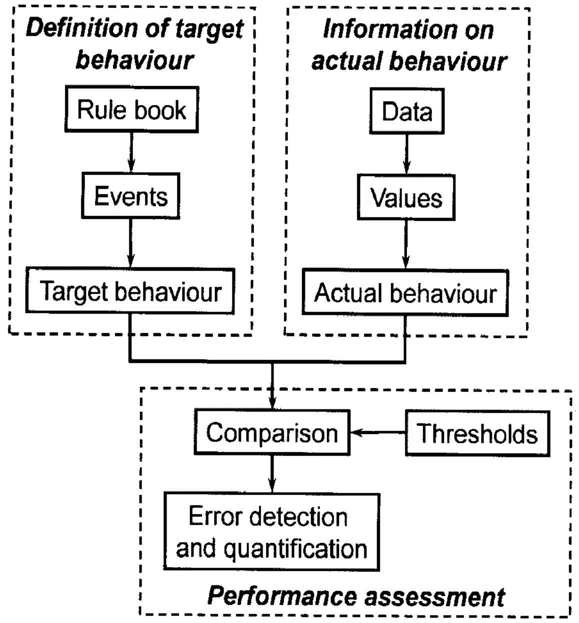

Fig. 1 PERMA Concept for the Assessment of Train Drivers by using a Simulator

2.2 Target behaviour, actual behaviour, and performance assessment

For the assessment of the train driver's performance during a simulation exercise, a concept is needed that integrates the necessary data, its processing, analysis, and interpretation. On the one hand, this data concept describes the target behaviour during specific simulator events. On the other hand, the actual behaviour of the train driver is recorded. Finally, the actual behaviour is compared with the target behaviour using thresholds for performance assessment. The PERMA concept (PERformance MArker) of assessment consists of the following elements (see Fig. 1):

1. Definition of target behaviour:

• Compilation of the rules and regulations that are relevant for running a train in a specific railway system (e.g. rule book, directives, speed book)

• Selection of training situations the train driver has to deal with (events)

• Derivation of the behaviour from the rules and regulations that has to be carried out by the train driver during an event (target behaviour)

2. Information on actual behaviour:

• Provision of the variables that reflect the driving behaviour of the train driver (data)

• Extraction and definition of the output values that describe the driving behaviour in specific events (values)

• Description of the behaviour of the train driver during an event (actual behaviour)

3. Performance assessment:

• Definition of criteria for the rating of deviations between target and actual behaviour (thresholds)

• Comparison of the derived target behaviour with the recorded actual behaviour and application of the thresholds

• Description and grading of the performance of the train driver (error detection and error quantification)

The basic procedure of the comparison is generally valid, but its application is unique for every event/situation of the simulator exercise. The relevant information for the assessment is provided by a data interface linked to the simulator (objective assessment of actual behaviour), the observation by an instructor (subjective assessment of actual behaviour), and an expert system that is filled with the ideal behaviour (target behaviour and thresholds). The assessment system has to be connected in run-time to the simulator and receives the information about the actual behaviour of the driver, i.e. actions on the controls, the status of the train (speeds, door status, etc), and the conditions of the surrounding environment (like route, signalling, etc.) through the data interface. The target behaviour of every single action is pre-defined and stored in an expert system. The data of both information sources–data interface and expert system–enter together into an assessment unit that is responsible for the assessment procedure itself.

When the train drives under normal, irregular or abnormal conditions, the driver always has to respect the operational rules and regulations. The assessment unit proves whether the train driver executes all necessary actions in accordance with the rules, i.e. in accurate order, in time and precise enough, and if deviations between actual and target behaviour occur. In addition, there is a need to define how these deviations should be rated in regards to the criteria safety, punctuality, and economy. Deviations from the target behaviour represent a poorer performance. For the assessment of the trainee, it is crucial to weight possible deviations from the target behaviour. If the mandatory behaviour is highly relevant for safety, it has to be weighted seriously. If the ideal behaviour is more a 'should' than a 'must' an advice to the trainee may be enough as feedback. If thresholds are defined for the assessment values, different error levels for the actual behaviour of the train driver can be allocated (e.g. advice, medium error, or safety-relevant error). Often, the thresholds have to be specified by an expert group as the rules and regulations of the train operating companies give no explicit threshold values (e.g. for speeding).

3. PERMA process model for driver performance assessment

The PERMA model (Fig. 2) outlines the steps that have

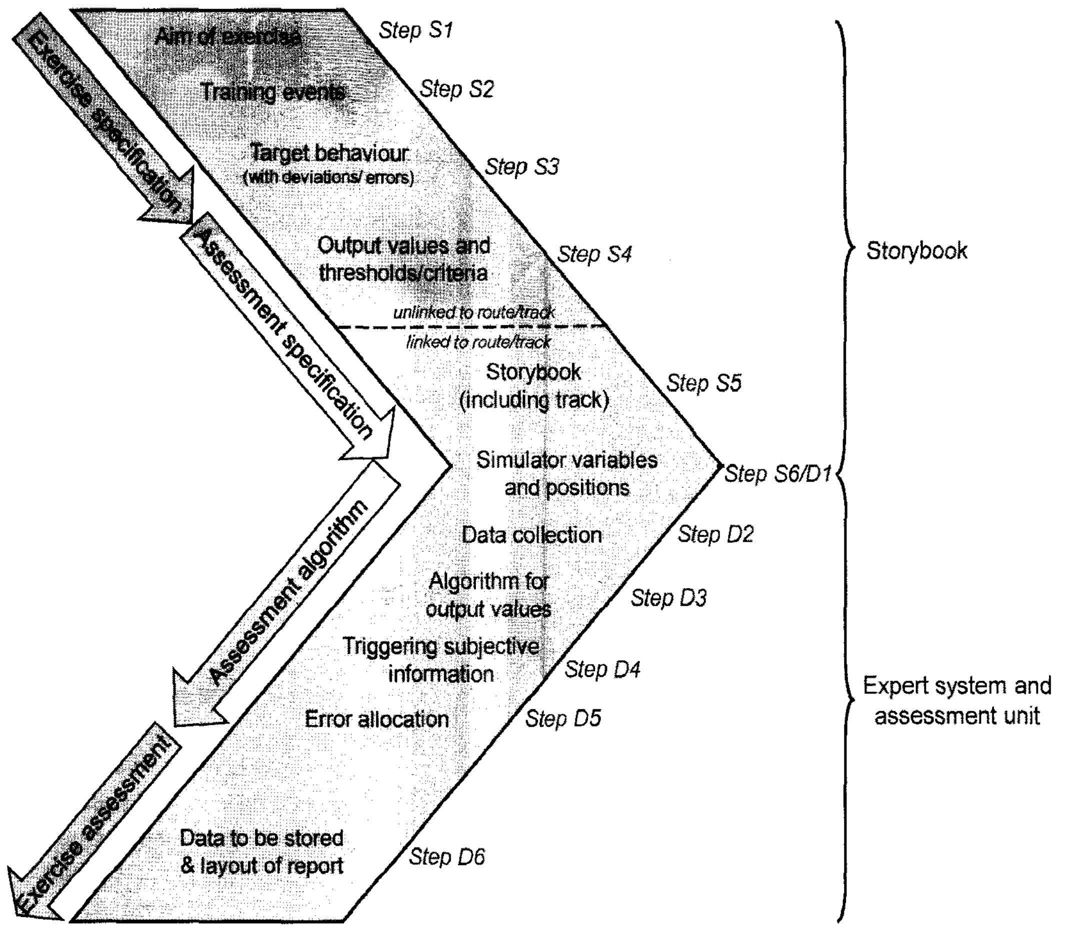

Fig. 2 PERMA model with preparative tasks done during the writing of the storybook and the rule development for the expert system and assessment unit

to be taken to implement the PERMA concept for the assessment of train drivers:

3.1 Description of the process model

The PERMA model (Fig. 2) illustrates the preparative tasks by splitting them in two different major components: (1) Storybook and (2) Expert system & assessment unit. The storybook specifies the exercise and defines the target behaviour and the assessment parameters (steps S1 to S6). The expert system and assessment unit has to be fed with the information necessary for assessing the exercise, e.g. relevant simulator data, assessment parameters/ criteria, feedback for the driver, assessment report, and data to be stored (steps D1 to D6). The preparative tasks have to be done for each simulator exercise. This significant effort is the prerequisite for an objective and transparent performance assessment.

The PERMA model starts top left with general considerations about the training aim of the simulator exercise (S1). By moving to the right side the specification in the storybook becomes more and more detailed. At first, the events that are embedded in the exercise are selected (S2). After that, the target behaviour of the train driver (as stated in the rule-book) is written down (S3). This step includes the specification of the error allocation for potential deviations. As a fourth step, the assessment specification starts with the specification of the assessment values (applied on simulator data) and subjective criteria (regarding the observation by an instructor) (S4). Whereas the steps S1 to S4 do not involve the actual route/track on which the exercise is driven, the next step specifies the concrete simulator scenario including the track/route, stations, signals, etc. (S5). This step links the training events with the route, e.g. the particular signal that leads to an automatic application of the brakes is chosen from the list of all signals available. The information on the output values and thresholds/ criteria together with the actual track specifies the simulator variables that are necessary for assessing the performance during the simulator exercise (S6).

The specified simulator variables are also the starting point for the definition of the assessment algorithm in the expert system that is used by the assessment unit (D1).

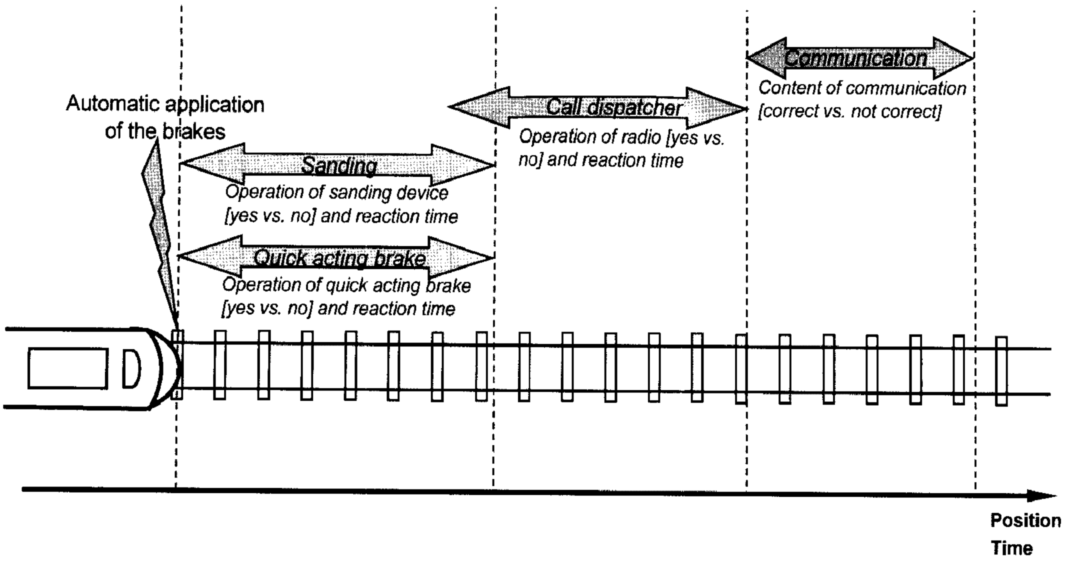

Fig. 3 Target behaviour after an automatic application of the brakes

From this very detailed level on the right the process becomes more general within the next steps. Based on simulator variables and positions the collection of data is defined (D2). The third step defines the algorithms for the calculation of the output values (D3). The fourth step defines the triggers for asking the real instructor for subjective information (D4). Afterwards, the error allocation is defined, i.e. which deviations result in which errors (D5). The definition of the content and layout of the assessment report completes the list of preparative tasks (D6).

3.2 Application of the PERMA model

A former national project has shown the applicability of the PERMA concept and its appropriateness for efficient research and training purposes. In the following, the application of the PERMA concept and model is illustrated by an example based on the regulations of a national train operating company (Fig. 3):

- Step S1: At the beginning the aim of the exercise is defined. For example: Simulator training can be done in order to check the competence of the train driver.
- Step S2: An event is a specific operational or technical occurrence in the course of a simulator exercise. An exemplary event during a simulation exercise could be that a train receives an automatic application of the brakes.
- Step S3: The rule-book sections define the target behaviour for the event of an automatic application of the brakes: The train driver has to support the train stopping by an application of the quick-acting brake and by operating the sanding device. Then, the driver has to call and inform the signaller about the exact position of the train and the reason for the automatic application of the brakes.
- Step S4: If the actual behaviour differs from the target behaviour, thresholds and criteria decide whether a deviation represents an error or not and whether it leads to an advice, medium or safety-relevant error.
- Step S5: For the implementation of a simulator exercise that consists of several events a detailed storybook is necessary. The storybook introduces the specific route/track (e.g. from station A to station B) on which the exercise is carried out.
- Step S6/D1: By a data interface a huge number of driving variables is available (e.g. positions of control levers in the cab like brake handle, sanding device, and radio). Some performance areas of the train driver cannot be assessed by the analysis of simulator data and therefore have to be assessed by a human observer, e.g. the content of the radio communication.
- Step D2: The data needed for the assessment are collected. For the exemplary event, this data include the exact onset time of the indicator that informs the train driver about the automatic application of the brakes, the exact points in time when the driver puts the brake handle and sanding device in the correct position, and the point in time when the train driver starts the radio call. For the content of the communication some questions to the human instructor provide the relevant information.
- Step D3: Output values that are calculated on the basis of the simulator data give information about the performance during the specific event, e.g. reaction times. These values describe the actual behaviour of the train driver during the event. The necessary output values calculated are the periods of time between the onset of the indicator for the automatic application of the brakes and (a) the start of the quick acting brake, (b) the start of sanding, and (c) the start of the radio call.

• Step D4: The conditions for requesting subjective assessments are specified. After the train has stopped, the human instructor is prompted to answer the question, if the content of the communication is correct or not.

• Step D5: The thresholds for assessing the deviation between actual and target behaviour are defined in detail. Thresholds specify significant deviations. For example: If the radio call is done later than one minute after the train stops, an advice should be printed in the final assessment report.

• Step D6: The layout and content of the assessment report are specified (general information concerning person, scenario, and train, graphical overview of the exercise, objective assessment values, subjective ratings by the human instructor, comments).

4. Description of the 2TRAIN add-on tools

After describing the assessment concept and the process model this section lines out the functionalities of the different software tools that enable the realisation of the PERMA concept at existing simulators. The first step is the development and implementation of a common data simulation interface (CDSI) in existing driving simulators of different simulator manufacturers. Every further simulator technology can be built on this interface and new technological developments can easily be adopted. In order to realise a reasonable performance assessment, the simulator has to compare the actual trainee's behaviour with valid rules and regulations. For this task, the regulations and directives are stored in a rule-based expert system (ExSys). The ExSys consists of two software modules: The rule editor and the online graphical user interface (GUI). The rule editor stores the target behaviour for different operational situations – an important precondition for a standardised and objective assessment of the driver's performance. To improve this procedure, specific error indices for the actual behaviour can also be allocated based on individualised assessment thresholds and parameters. The online GUI displays the whole assessment procedure during the simulation run and allows the human instructor to enter additional ratings and further comments. The online assessment is done by a virtual instructor (VI) that has access to all simulator data, especially those that are difficult or even impossible for the human instructor to observe. Assessing simulation data in real time also gives the option to support the driver during and guide a trainee through an exercise by means of adaptive training or enhanced reality. The final results are stored in an assessment database (AssDB) permitting the training

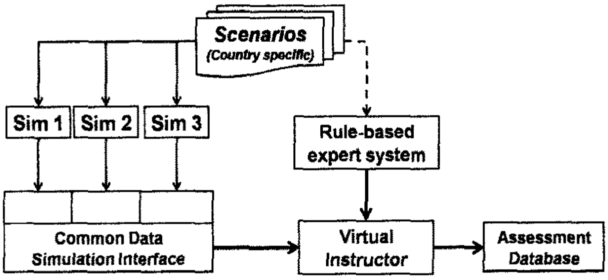

Fig. 4 Illustration of the extended simulation system

department to monitor the individual learning history and the competence level of a given trainee. Additional data deriving from information sources like CBT/WBT or real vehicle data can serve as further input. Fig. 4 gives an illustration of the whole system.

4.1 Common data simulation interface

The CDSI guarantees the implementation of the 2TRAIN tools at different simulators developed by different manufacturers. The interface provides a set of variables that are necessary for the assessment of train drivers' performance in given simulation scenarios. As the regulations and directives of these scenarios are country specific, the list of variables cannot be common and has to be adjusted for each simulator.

4.2 Rule-based expert system

The ExSys provides the target behaviour of the train driver for particular scenarios in the form of rule sets. A rule describes how to test the actual behaviour of the driver against the target behaviour for a specific training event and is closely related to the regulations. Each rule defines and uses assessment parameters, thresholds, and tolerances and each rule set is compiled for a particular simulation exercise. The rule set is then used by the VI to execute the actual assessment of the simulation run. The ExSys consists of two modules: The first one is the ExSys rule editor (Fig. 5) that allows the creation, adjustment, and management of the rule sets.

The ExSys rule editor supports the construction of rule templates that transfer the regulations and directives into a computer-readable format. The rule-template browser compiles a set of rule templates that are needed for the assessment of different simulator events. Then, the rule-instances browser edits the values of the assessment parameters, i.e. instantiates the rule templates with the parameters of the actual simulated track and the actual exercise. A simplified schema of the track is displayed at the right-hand side of the editor in order to support the compilation and instantiation of the rule set. Presumably, the value of the parameters will be changed more often

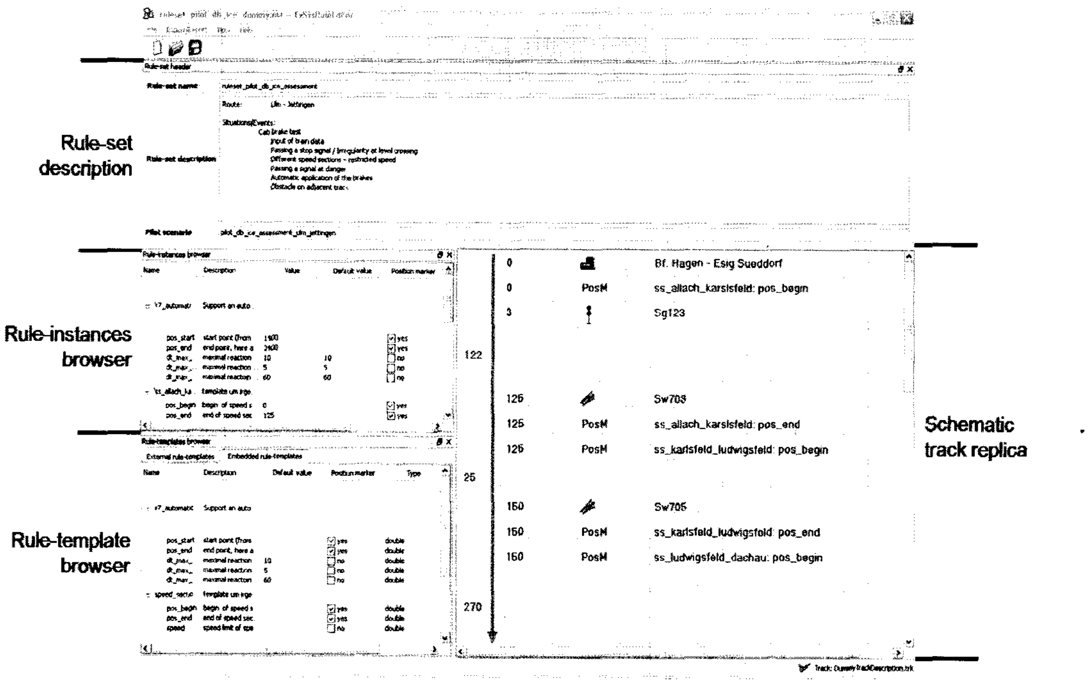

Fig. 5 Structure of the ExSys rule editor

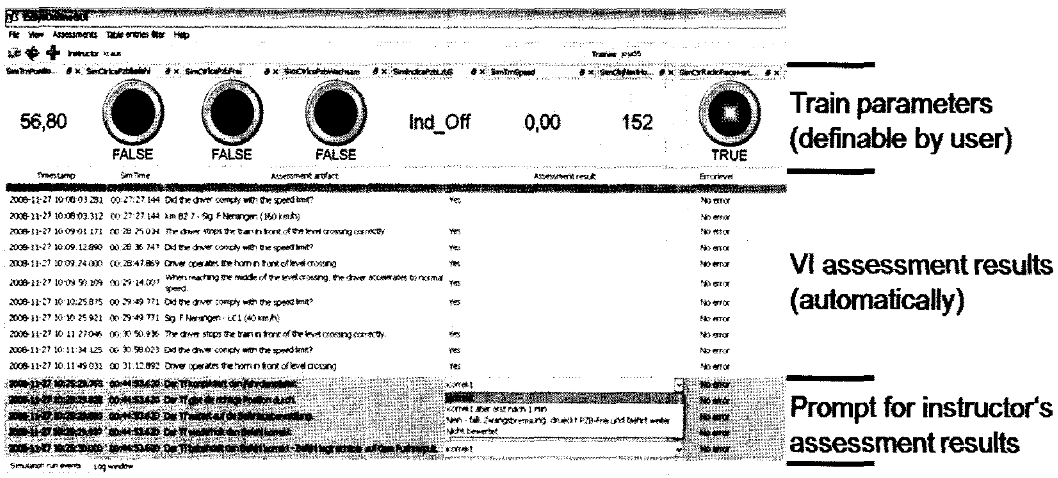

Fig. 6 Structure of the ExSys online GUI

than the rule templates themselves. Probably, some of the rule templates can be used in more than one country, but have to be configured differently. After its finalisation, the completed rule set can be provided for the online assessment done by the VI. The rule editor is able to run standalone without connection to a pilot simulator, as no particular input from the simulator is needed.

The second ExSys module is the ExSys online GUI (Fig. 6) that provides an interface for the human instructor to monitor and influence the assessment process during the simulation run. All assessment results that are based on the automatic assessment are displayed automatically. But not all aspects of the driver performance can be handled by the VI. For example, the correct content of a radio message like an emergency call can not be rated by the VI. The assessment of this kind of situations is in charge of the human instructor. In these cases, the VI triggers an accordant row and requests the instructor to enter the assessment result.

4.3 Virtual instructor

The VI phrases and evaluates a set of rules in real time given by the ExSys. According to these rules the VI compares the actual and the target behaviour, sends the results to the ExSys online GUI, and finally creates a simulation report which is sent to and stored in the AssDB when the exercise ends. This *assessment mode* of the VI runs in the background, i.e. is invisible for the trainee, and is mainly

Table 1 Recommended application areas of the VI working modes

	Assessment mode	Help mode	Guidance mode
Purpose of training	refresher training, competence checks	acquisition of new contents, repetition	acquisition of new contents
Type of simulator	full cab, partial cab	all types	partial cab, part-task trainer, software interface
Ratio of trainer/trainee	one-to-one	one-to-one, one trainer/more trainees, no trainer present	one trainer/more trainees, no trainer present

designed for advanced trainings or competence checks. As the trainee gets no feedback during the exercise, it is important to provide a detailed debriefing with the human instructor explaining the simulation report at the end of the simulation run.

As the VI works online, it is also possible to give direct feedback to the trainee in case of a deviation from the target behaviour. This *help mode* can be simply designed as short text messages or icons that are displayed in the visual system of the simulator. It may also be possible - depending on the technical characteristics of the used simulation system - to enhance the simulated reality, e.g. spotlight important lineside information like relevant signal aspects or speed tables. The VI may also initiate a repetition of an operational situation if something went wrong or skip to a more difficult and advanced situation. This adaptive mode supports the individual competence level and increases the individual learning gain. The help mode is mainly designed for train drivers in education and therefore for initial training.

The third working mode is called *guidance mode* and combines the simulation exercise with (1) integrated learning modules and/or (2) advisory details about the simulator session. Concerning the first point, there are different alternatives how to create this information. One possibility is a simulation embedded learning module similar to a short CBT section. This learning module explains the upcoming operational situation and the necessary and required actions. Once the exercise is completed the VI freezes the simulator and starts a new learning module that provides explanations concerning the next situation of the simulator exercise. The second point contains information about how to start the session, how to operate the simulator, and about the contents and structure of the session. These advisory details support self-paced simulation training and are – due to a missing human trainer – not applicable for all types of simulators. The learning modules and/ or advisory details are presented on an additional screen nearby the driver desk or – in case of a desktop simulator – on the screen where the simulation runs. In order to keep the simulation session realistic, they require only a limited interaction with the driver.

Whereas the help and guidance modes are especially designed for training purposes, the assessment mode is also applicable for research questions related to the driving task of a train driver. The three modes of the VI should be understood as additional supportive functions designed for different stages of training. The recommended usage of these different modes depends on the purpose of training and the competence level of the trainee (acquisition of new contents, repetition, refresher training, and competence check), the type of simulator (full cab, partial cab, part-task trainer, software interface), and - related to both points above – the ratio of trainer and trainee (one-to-one, one trainer/more trainees, no trainer present) (see Table 1). For example: As the assessment mode provides no feedback to the trainee, a briefing and/or debriefing is essential to enhance the learning gain. In addition and due to safety reasons, a training session on a full cab simulator should always be accompanied by an instructor. But simulators – in particular software interfaces – can also be used for self-paced learning units where no trainer is needed.

In summary, the main tasks of the VI are: start and stop the assessment rules, evaluate active rules in real time, send help messages and guidance lessons to the driver, stop, play and resume an exercise to give additional information to the driver, trigger and receive subjective assessments from the instructor, receive comments from the instructor, and send a log file to the AssDB once the exercise has been finished.

4.4 Assessment database

The AssDB stores the assessment data of every trainee after the simulation run and creates the simulation report. Once, the results of a significant sample have been stored in the database, it is also possible to provide relative assessment by comparing an individual datum with grouped data. This also enables the system to run this comparison online during the simulator session, in order to enhance the validity of the result. For example, the statement that the reaction time of the individual trainee in a given situation is better than that of 75 percent of all trainees makes more sense than the blank result that the reaction time was 4.2 seconds.

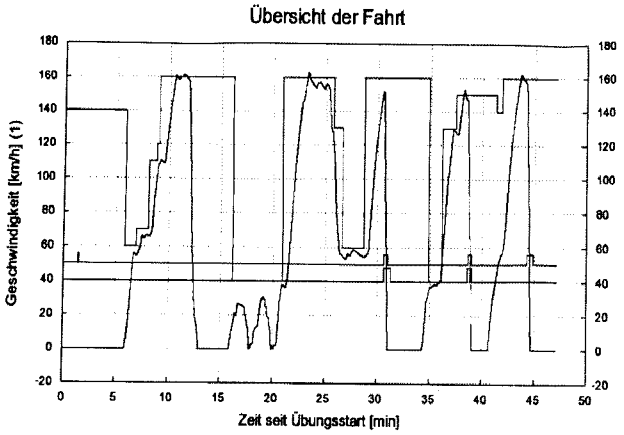

Fig. 7 Illustration of a graphical overview of the exercise

Did the driver comply with the speed limit?	No, more than 5 km/h faster
LC2 km 73.8 - Sig. P1 Nersingen (40 km/h)	Speed limit: 40.00 km/h Max. speed: 49.21 km/h (AFB)
	Time > speed limit: 9.34 s
Irregularity at level crossing	
The driver stops the train in front of the level crossing correctly.	Yes
Did the driver comply with the speed limit?	Yes
Driver operates the horn in front of level crossing	No, too short (< 2s)
	Duration: 1.5 s Speed: 0 km/h
When reaching the middle of the level crossing, the driver accelerates to normal speed.	Yes

Fig. 8 Example of a report section (speed section/specific scenario event)

This database should not only store the simulation results but should function as a central learning and training register where all the training data and results are stored. If this database is embedded in the overall competence management system, it allows not only a long term learning data recording and an analysis of an individual trainee or a group of trainees but also provides implications about frequent errors made during specific training courses.

4.5 Simulation report

The simulation report summarises the exercise results. A summary of the critical errors made during the simulation run is displayed at the top of the report, in order to see the most important result: Has a safety-critical mistake occurred? A graphical overview displays assessment parameters of interest. The parameters that the graph contains can be configured by the user. In Fig. 7, the rectangular line represents the target speed and the curved line represents the actual speed. Whenever the curved line is above the rectangular one, the trainee has exceeded the speed limit. This overview is a supportive tool for the debriefing of the session.

The main part of the simulation report is a chronological order of each relevant situation throughout the simulator run. In case of a deviation the report contains more details than the information given via the ExSys online GUI. For example, if the speed limit was exceeded, the report displays the speed limit, the actual maximum speed driven, and the duration of the speed exceeding (see Fig. 8).

5. Demonstration activities and evaluation results

The add-on tools were implemented amongst other simulators at a simulator training centre in Germany. 44 train drivers from different divisions and undertakings ran through a four hour training, in order to evaluate the reliability and acceptance of each tool. Some results that are directly linked to the add-on tools are:

- The instructors pointed out the supportive character of the ExSys online GUI during an exercise. It helps to focus on the observation and to concentrate on the communication processes as part of a scenario.
- A throughout positive response is given for the help mode of the VI. The feedback of the system in case of a deviation makes it possible to directly adjust the behaviour what - as a consequence - enhances the learning gain.
- Both, instructors and train drivers, appreciate the usage of driving and behaviour data (e.g. maximum speed, reaction time) in order to improve the objectivity and quality of performance assessment. As far as the storage of the results in the AssDB is concerned the trainees are sceptical due to the comprehensive character of such a database.
- The results from the demonstration activities show that the assessment report is a proper method to support the debriefing. It highlights the lack of performance and helps to argue and to give proof of failures made by the trainee during the exercise.

6. Conclusion and recommendations

Using the described PERMA concept and process model for the assessment of the train drivers' performance during a simulation exercise whether for training or research purposes has several benefits:

The performance can be measured in a wide range of realistic situations that reflect the complexity of the train driver's task. These situations include degraded and abnormal operational conditions. Furthermore, events can be realised that cannot be reproduced in reality (e.g. equipment failures) and are very rare during daily operation (out-of-course events). Additionally, the events used for assessment can be reproduced in a repeatable, controllable and consistent way. This allows a standardised and objective assessment as every tested train driver has to undergo an identical testing scenario. This is a vital necessity for research but also an advantage for training.

The availability of performance recordings and assessment facilities ensures a (semi-) automated and objective comparison of the driver's performance against pre-set standards and criteria. For the overall assessment, the objective data from the simulation can be complemented with subjective ratings by the instructor. In a training context, analysing the assessment data is an important basis for feedback and debriefing. Furthermore, it helps to improve the quality of the overall training and assessment programme and can be used to review the effectiveness of the whole training application.

A difficult point is the construction of a rule set that is readable by the VI. One reason is the sheer amount of regulations; another reason is the complexity as most of the regulations depend on a given operational situation. An aggravating factor is the ongoing process of changing and adapting regulations in the daily operation. But it is not necessary to transfer all regulations for every conceivable situation. Former research has shown that a clustering of target behaviour in so-called behaviour classes is a promising way to simplify this task (Maag, Schmitz, Siebers & Krüger, 2004). Another possibility is the creation of rule templates. For example: It is possible to create a rule template that monitors the speed behaviour. A single rule template can then cover a whole range of speed regulations. The only input needed to instantiate the rule is the speed limit and a start and end point of a given speed section.

A pilot study with train drivers and trainers from Germany could show that the acceptance of the tools is rather good. Especially the help and guidance mode of the VI as well as simulation report were rated high by the trainees. The comprehensive storage of data is seen with more scepticism. The trainers commended the supportive character of the ExSys GUI and also the detailed simulation report. Two more pilot studies are planned in France and Spain. All evaluation results will be summarised in a best-practise guideline at the end of the 2TRAIN project.

Reference

1. Bommer, W. H., Jonson, J. L., Rich, G. A., Podsakoff, P. M. and Mackenzie, S. B.(1995), On the interchangeability of objective and subjective measures of employee performance. *Personnel Psychology, 48*, 587-600.
2. Dvorak, H. & Brunner, A.(1987), Computergestützte Diagnostik bei Vorsorgeuntersuchungen für Fahr-, Steuer- und Überwachungstätigkeiten, *Arbeitsmedizin Sozialmedizin Präventivmedizin, 22(9)*, 217-221.
3. Elms, D.(2001), Rail safety, *Reliability Engineering and System Safety, 74*, 291-297.
4. European Commission(2007), Directive 2007/59/EC of the European Parliament and of the Council of 23 October 2007 on the certification of train drivers operating locomotives and trains on the railway system in the Community, *Official Journal of the European Union, L315*, 51-78.
5. Groeger, J. A., Bradshaw, M. F., Everatt, J. & Field, D.(2001), *Pilot study of train-drivers' eye-movements.* Rail Safety and Standards Board: London.
6. Hall, S. & Brannick, M. T.(2009), Performance assessment in simulation. In D. A. Vincenzi, J. A. Wise, M. Mouloua & P. A. Hancock (eds.) *Human factors in simulation and training* (pp. 149-168). CRC Press: Boca Raton.
7. Maag, C., Schmitz, M. & Fröschl, T.(2009), Psychologie des Eisenbahnverkehrs. In H.-P. Krüger (Hrg.), *Enzyklopädie der Psychologie Anwendungsfelder der Verkehrspsychologie–Band 2* (pp. 639-709). Hogrefe: Göttingen.
8. Maag, C., Schmitz, M., Siebers, A. & Krüger, H.-P.(2004), *Development of performance markers for safety trainings in train simulations.* Seminar and Workshop for the Nordic Rail Sector, May 27th/28th, Fulda.
9. Maag, C., Schmitz, M., Siebers, A. & Krüger, H.-P.(2005), *Development of performance markers for safety trainings in train simulation.* International Congress Transport Safety, March 16th/17th, Fulda.
10. Oed, R. & Müller, B.(1995), Simulationssystem zur Erprobung der Mensch-Maschine-Schnittstelle im Führerstand. In VDI-Gesellschaft (Hrg.), *Simulation und Simulatoren für den Schienenverkehr* (pp. 61-70). VDI: Düsseldorf.
11. Rail Safety & Standards Board(2007), Good practise on simulation as a tool for training and assessment. Railway Group Standard RS/501, Issue 2, June 2007.
12. Schmitz, M. & Endres, S.(2008), *Unlock the potential of your simulator through a standardised approach to technology-based training.* Transport IQ's 7th Annual "Best Practice for Driver Management", February 26th/27th, London.
13. Schmitz, M. & Maag, C. (eds.)(2008), *Benchmarking report on computer-based railway training in Europe.* IZVW: Wuerzburg.
14. Schmitz, M., Maag, C. & Mera, J. M.(2008), The use of computer-based training tools in Europe - An overview and new approaches. In J. Allen, E. Arias, C. A. Brebbia, C. J. Goodman, A. F. Rumsey, G. Sciutto & N. Tomii (eds.) *Computer in Railways XI* (pp. 829-838). WIT Press: Southampton.