Transportation Research Part F: Psychology and Behaviour 97 (2023) 109–122

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

journal homepage: www.elsevier.com/locate/trf

A validation study comparing performance in a low-fidelity train-driving simulator with actual train driving performance

Niklas Olsson

The Swedish National Road and Transport Research Institute (VTI) and Luleå University of Technology (LTU), Sweden

ARTICLE INFO

Keywords:
Relative validity
COVID-19
Ecological validity
Simulator
Low-fidelity simulator
Train driver

ABSTRACT

Background: Although common in other industries, such as the automotive sector, no train-driving validation study has been found in the existing literature. The present paper intends to fill that gap by comparing the results of train-driving performance in a physically low-fidelity but highly functional simulator with real train-driving performance.
Method: Thirty-four train driver students in the final part of their basic education were assessed in a 45-minute simulator test using the number of driving errors as the performance indicator. The results were compared with the performance at 11 weeks of internship as measured by supervisors grading according to a standard procedure. One of the classes (17 to-be drivers) was affected by restrictions related to COVID-19, which led to a shortened internship and distance learning during parts of the internship. The study also intended to measure the effect of the restrictions and the types of errors the drivers made by comparing the two classes.
Results: A significant correlation was found between the number of driving errors and internship grades, $r = -0.45$, $p < .05$. The results also revealed that COVID-19 restrictions negatively affected performance, as the students from Class B made significantly more driving errors and obtained a lower internship grade than those from Class A.
Conclusions: This paper shows that this type of low-fidelity simulator is well suited for measuring real train-driving performance. A measurement method that can predict long-term driving should have implications for both research and practical usability. Researchers can use this for studying the effects of, for example, different training methods, while train operation companies can use the method to test their drivers' skills and intervene before an actual accident occurs.

1. Introduction

A driving simulator offers a safe and useful environment for cost-effective, controlled and standardised tests of human behaviour (Blana, 1996; Godley, Triggs, & Fildes, 2002). Due to railway capacity shortages and the difficulty of controlling real traffic situations, there is a need to conduct research on train-driver behaviour in a safe, simulated environment where the results are transferable to real train driving. In closely related industries, such as the automotive industry, such validation studies are common (Mullen, Charlton, Devlin, & Bedard, 2011; Wynne, Beanland, & Salmon, 2019); however, similar studies in railway simulators have yet to be identified. A valid measurement method for simulators would be beneficial for research, for instance, to evaluate various training methods, examine stress handling or examine the effect of experience among train drivers. In practice, a realistic, valid, and reliable test method that includes a sample of critical situations would help train operating companies (TOCs) examine to-be train drivers and assess whether

E-mail address: niklas.olsson@vti.se.

https://doi.org/10.1016/j.trf.2023.07.007
Received 1 February 2023; Received in revised form 5 July 2023; Accepted 10 July 2023
Available online 14 July 2023
1369-8478/

extra training is necessary to reduce the risk of an actual real-world accident. This paper intends to test the ecological validity of a low-fidelity simulator (LFS) measurement method to be used by researchers and by TOCs.

In this context, fidelity describes the extent to which the simulator replicates the actual system (Roberts, Stanton, Plant, Fay, & Pope, 2020). *Physical fidelity* is usually referred to when simulator fidelity is described, which describes how it feels, looks, and sounds physically (Hays, 1980). Wynne et al. (2019) defined the extreme of an LFS as a fixed-base simulator with a single computer screen controlled by the simplest control. In comparison, a high-fidelity simulator (HFS) is a fully equipped, full-motion cab with large screens and a 360° field of view (Wynne et al., 2019; Kaptein, Theeuwes, & van der Horst, 1996). The relationship between physical fidelity and the training effect has been debated (Dahlstrom, Dekker, van Winsen, & Nyce, 2009) but generally the level of fidelity has been considered less important when training general special cases and for novices (Saus, Johnsen, & Eid, 2010; Noble, 2002). However, high fidelity has been considered more important for experienced drivers and when practicing specific tasks with a particular vehicle to avoid mislearning (Noble, 2002; Myers, Starr, & Mullins, 2018). Another subcategory of fidelity is *functional fidelity*, which refers to a simulator's ability to incorporate all the same functions the same as in the real-world environment (Roberts et al., 2020). For example, if the simulated train decelerates, accelerates, or reacts to a maneuvered lever or a pressed button in the same manner as it would in reality, then the level of functional fidelity is high.

Functional fidelity in the simulator is considered as, or even more, important than physical fidelity with regard to transferability to real-life performance (Dahlstrom et al., 2009; Roberts et al., 2020; Hamstra, Brydges, Hatala, Zendejas, & Cook, 2014).

Starting a few years ago, a simulator with low physical fidelity but a high level of functional fidelity has been widely used among Swedish train-driver educators and operators (Thorslund et al., 2019) within, e.g., basic education, annual learning, and examinations. In the present study, the performance of 34 train-driver students in such a simulator is compared to internship performance.

A simulator's ability to replicate human behaviour, in reality, is described as behavioural validity (Mullen, et al., 2011) and can be divided into two subcategories: *absolute behavioural validity* prevails when the values obtained in a simulator experiment are identical or almost identical to the values in reality, while *relative validity* means that the effects from a simulator study point in the same direction as the results shown in the real world but do not match in absolute numbers (Blaauw, 1982). Driving simulator studies commonly aim for the latter, and almost exclusively so when an LFS is used and has been investigated via various measurement methods of driving performance (Wynne et al., 2019; Mullen et al., 2011).

A relationship between simulator performance and real-life indicators other than actual driving performance has been found in some studies, such as with a neuropsychological test that was used to predict safe driving (Bedard, Parkkari, Weaver, Riendeau, & Dahlquist, 2010); the risk of making traffic violations (Hoffman & McDowd, 2010); or by showing that experienced drivers perform better than novices in a simulator test (as they are assumed to also do in reality, see Dorn & Barker, 2005). However, the most common measurement method aiming for relative validity is to compare performance between real driving and simulator driving (Galski, Bruno, & Ehle, 1992; de Winter et al., 2009; Mayhew, Simpson, Wood, Lonero, Clinton, & Johnson, A, 2011; Lew et al., 2005). When using LFSs, a broad perspective of performance is often measured in the same test (i.e., overall performance), for example, speed handling, positioning, rule handling, efficiency, reaction time, driver workload, and communication management (Aksan et al., 2016; Lee, Cameron, & Lee, 2003; Freund, Gravenstein, Ferris, & Shaheen, 2002; Lobjois, Faure, Désiré, & Benguigui, 2021; Tichon & Wallis, 2010).

When the results of a study can be generalised to the real world during an extended, uncontrolled period of time, the results are said to be *ecologically valid* (Schmuckler, 2001), which is rare among driving-simulator studies, although some examples from the automotive industry exist. For example, Lew et al. (2005) used family members to evaluate a driver with a brain injury during a stretch of four weeks using a wide variety of performance indicators, while Hoffman and McDowd (2010) and Edwards, Hanh, and Fleischman (1977) found validity by correlating involvement in accidents over a longer period with performance in a simulator environment.

1.1. *The train-driver profession*

Previous studies have described train driving as complex, where both lineside information (e.g., signals, marker boards, topography, and weather) and in-cab information (i.e., data from the train and the train protection system) need to be processed and handled according to the regulations to accomplish safe and efficient driving (Branton, 1979; Naweed, 2014). The development of technical aids has reduced the number of accidents (Evans, 2011) and changed the cognitive tasks of train driving, such that the in-cab information can be seen as an additional source of information to consider (Hamilton & Clarke, 2005; Naweed, 2014; Buksh, Sharples, Wilson, Coplestone, & Morrisroe, 2013). Nevertheless, in the 2020s, many tracks are not equipped with technical systems, and when available, they can fail, and such a situation must be handled safely and efficiently (Naweed, 2014; Harms & Fredén, 1996; Jansson, Olsson, & Kecklund, 2000). Furthermore, the driver needs to be aware of additional information, such as the route book, the timetable, the rulebook, a tablet with digital information and, finally, different kinds of safety messages from the signaller (Forsberg, 2016) (Kecklund et al., 2001).

A comprehensive Swedish study of accident reports between 1980 and 1997 found that a *deviation from the normal process* often occurred prior to an accident (Kecklund, Ingre, Kecklund, Söderström, Lindberg, Jansson, Olsson, & Sandblad, 2001). To avoid being involved in accidents, numerous studies highlight sustained attention as one of the most important abilities (Baysari, Caponecchia, McIntosh, & Wilson, 2009; Edkins & Pollock, 1997; Tabai, Bagheri, Sadeghi-Firoozabadi, & Shahidi, 2017). The reasons for failed attention are categorised as *distractions*, referring to something that causes the driver to shift attention from the primary task, such as a sight restriction, another person in the cabin or a phone call with a signaller (Phillips & Sagberg, 2014; Verstappen, 2017; Naweed, 2013); *conflicting goals* (e.g., time pressure that causes the driver to compromise safety, see Hickey & Collins, 2017); *monotonous driving* and *fatigue* (Dunn & Williamson, 2012; Dorrian, Roach, Fletcher, & Dawson, 2007; Fan et al., 2022). Creating an accurate situational mental model based on the information from different sources (i.e., to be situationally aware) is also vital for train drivers to avoid accidents (Roth & Multer, 2009; Luke, Brook-Carter, Parkes, Grimes, & Mills, 2006). Likewise, to avoid Signals Passed At Danger (SPADs), prospective memory (i.e., remembering what the next signal should be) is required (Luke et al., 2006; O'Connell, Lawton, Mills, & Klockner, 2017). Other studies also emphasise *communication misunderstanding* as a risk factor (Murphy, 2001; Shanahan, Gregory, Shannon, & Gibson, 2007; Smith, Kyriakidis, Majumdar, & Ochieng, 2013).

1.2. Simulators within train driving research

Several publications have been made as a result of controlled simulator experiments for many purposes. A couple of studies found a positive effect of repeated practice (Tichon & Wallis, 2010; Olsson, Lidestam, & Thorslund, 2022) but no effect of the level of simulator fidelity on performance (Tichon & Wallis, 2010). A couple of papers studied the effect of train-driving experience on, i.e., the driving-style (Large, Golightly, & Taylor, 2017) and attention (Du, Zhi, & He, 2022). Simulator experiments have displayed that performance was affected negatively through time-pressure (Suter & Stoller, 2014) and distraction (Verstappen, 2017). Several studies have shown that sleepiness and monotonous driving increase the risk of mistakes (Dorrian et al., 2007; Dunn & Williamson, 2012; Brandenburger & Jipp, 2017), and train driver vigilance has also been tested by using EEG signals collected from equipment placed on the driver (Fan et al., 2022; Zhang et al., 2017). In the last decade, studies of how different levels of technical aids affect driver workload have been common (Brandenburger, Hörmann, Stelling, & Naumann, 2017; van der Weide, 2017; Buksh, Sharples, Wilson, Coplestone, & Morrisroe, 2013).

The present study aims to compare the performance of train-driver students on a simulator test against performance in real-world driving during internships. The students consisted of two classes, of which one (Class B) was affected by COVID-19 restrictions, such that 10 out of 33 weeks of classroom learning were replaced by distance learning via online video conference software, and approximately two weeks' reduction of the internship. Half of the students in each class had a peer-friend watching from another room when the simulator test was conducted, and the peer-friend was instructed to give feedback to the driver afterward (this design was due to that the simulator test was also part of another study aiming to measure the effect of peer-feedback). This aimed to measure whether a peer was watching affected driving performance in any direction.

1.3. Aim and research questions

The aim of this paper is to compare train-driver students' performance in an LFS with performance from an eleven-week internship. The first three questions were primary to investigate while the last two were added due to external circumstances. The research questions were as follows.

(1) How does train-driver student performance in a simulator test correlate with internship performance?
(2) How do internship time and age correlate with performance from the simulator test and the internship, respectively?
(3) Which types of driving errors are made by student train drivers in a simulator test?
(4) How did the COVID-19 restrictions affect simulator-test performance, internship time, and internship performance?
(5) How does peer watching from another room affect simulator-test performance?

2. Materials and methods

2.1. Study design

For RQ1 and RQ2, a correlational design was used, comparing train-driving performance in a simulator test, internship performance, internship time, three types of driving errors and age.

For RQ3, a split-plot factorial design was used. The between-subject factor was Class (A; B), and the within-subjects factor was Error Type (rule-based errors; skill-based errors; ineffective handling errors).

For RQ4, the effects of COVID-19, a between-subjects design, were used, with Class (A; B) as an independent variable for comparisons on three separate dependent variables: simulator performance, internship time, and internship performance.

Finally, RQ5 was tested with a 2 × 2 between-subjects design: Class (A; B) × Peer Watching (with; without).

2.2. Participants and train-driver basic education

The sample was two classes and a total of 34 Swedish train-driver students (27 males, 7 females), aged from 21 to 50 years (M = 31.8, SD = 8.08 years). The two classes underwent the same selection process (including psychological tests and health screening), followed the same syllabus, and had, for the most part, the same teachers in the theoretical part of the education. However, Class A took place in 2019–2020 and Class B followed in 2020–2021. At the time of the study, they were heading into the final part of a 60-week train-driver education program, 22 weeks of which consisted of internships divided into 5 periods. The last 3 periods (11 weeks) were included in this study due to their temporal proximity to the simulator test. (Note: the two omitted internship periods took place long before the simulator test). The actual simulator test was conducted within the five weeks of classroom theory between internships 4 and 5. Due to restrictions from the COVID-19 pandemic, Class B experienced approximately two weeks less internship time, and 10 out of 33 weeks of classroom theory was replaced by distance education. Distance learning was conducted via online video conference software with live lectures where the teacher was mainly situated in the classroom and the students at home. It should be noted that this design was not part of the original plan nor research question. See Fig. 1 for the participants, syllabus, and study design details.

All the participants had, during their basic education, driven a similar train-driving simulator as the one used in the test but had not practised any of the special cases in this study.

2.3. Apparatus

The fixed-base portable simulator used in this study was developed by Swedish engineers at VTI (The Swedish National Road and Transport Research Institute) in close collaboration between researchers and end-users, always with the intention of reaching a high level of functional fidelity (Thorslund, Rosberg, Lindström, & Peters, 2019). The simulator consists of a touch-screen PC, a driving lever, and a braking lever connected to a 50-inch flat screen, approximately two meters in front of the driver, displaying the outside view (see Fig. 2). The graphics are built using the open-source software OpenSceneGraph. The outside rendered field of view in reality depends on the locomotive, but the main difference in the simulated view is the missing side windows (but, since a train cannot be steered, cannot brake within the line of sight, and nothing is to detect on the side of the train, the simulated field of view should be sufficient). The speedometer, train protection system, and some train-specific buttons could be seen and operated on the PC, while the airbrake manometer used for the freight-train braking system appeared on the flat screen. The levers used come from real trains, and the retardation and acceleration values inserted in the software correspond to real train driving, as do the values entered in the train protection system. The simulator includes two vehicle models, both of which are among Sweden's most common. For the freight train, a model in accordance with the Bombardier locomotive TRAXX was used (developed and validated within a project financed by the Swedish Transport Administration, see Andersson, Lidström, Peters, Rosberg, & Thorslund, 2017), and for the passenger train the model was based on the Bombardier Regina EMU (X50). The Swedish train protection system (called ATC) corresponds with real train driving in Sweden. According to the classification by Wynne et al. (2019), the simulator should be considered an LFS in physical terms but with high levels of functional fidelity.

The tracks used in the simulator were built as real tracks based on infrastructure data collected from BIS (the Transport Administration database), including coordinates, curvatures, signal positions, level-crossings, gradients, and balise positions. To mimic the track surroundings, video films from the routes were used. Thus, the participants in the study used the same document (route book) as

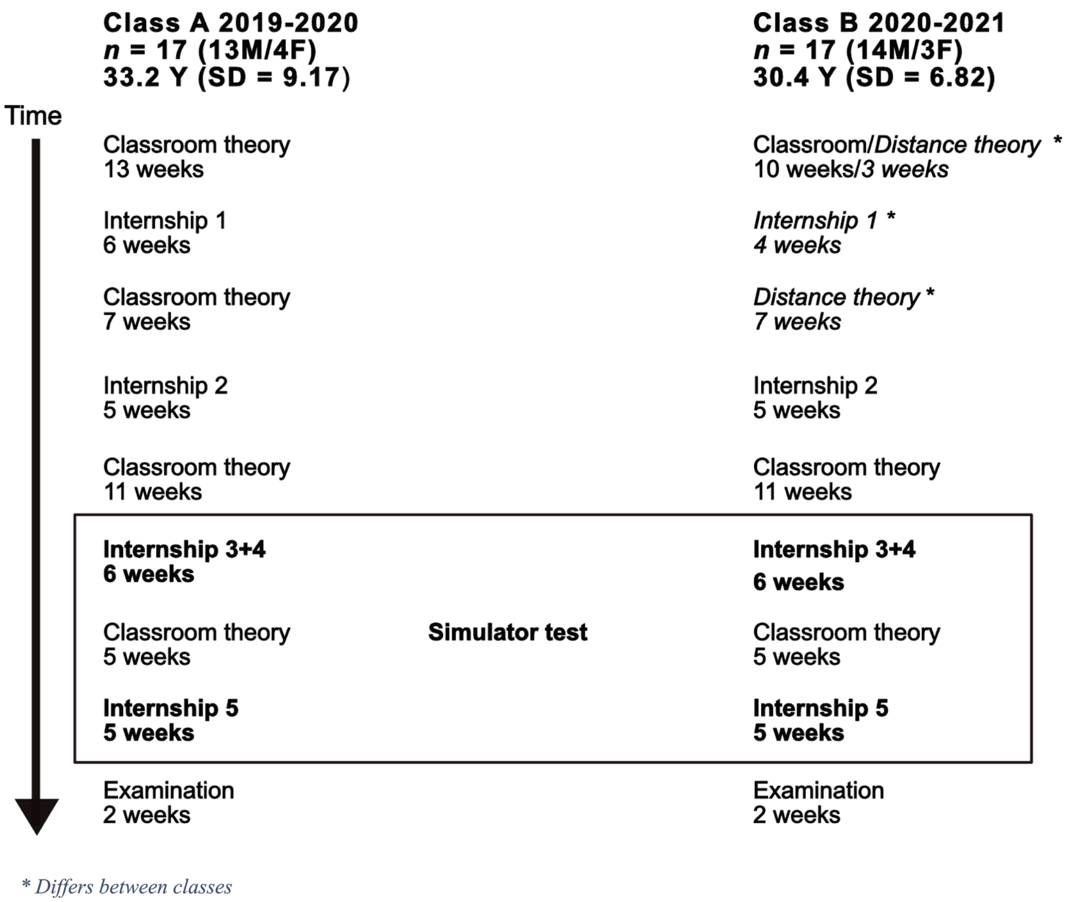


Fig. 1. Study design and basic train driver education.

Fig. 2. Simulator environment.

they would in reality, where they can see signals, signs, level crossings, stations and other details about specific locations. The software is continuously updated and validated per the need of the current 14 (as of 2023) Swedish TOCs and educators who use the simulator and as a result of different research projects. However, the users receive updated software twice a year.

During the study, an instructor acted as a signaller and interacted with the real writing forms and according to the regulations, although not by telephone calls as would have occurred in reality.

2.4. Simulator test procedure

Before the test, the drivers drove a freight train for approximately 15 min to get to know the simulator environment, its levers and buttons. Prior to the test, they also had a chance to study the documents used, which were the same as during their internship (i.e., timetable, safety forms, and line description).

The test consisted of two 20–25-minute scenarios, one of which ran a freight train (with airbrakes) and the other a passenger train (with electric-pneumatic braking system), with a short break in between. The drivers were alone with an instructor who acted as a signaller in the study.

The test was filmed and recorded to facilitate later assessment. Half of the participants (from each class) had a peer watching the simulator drive from another room with the instruction to give peer feedback after the study had ended. The driver was aware of this but could neither hear nor see the other student. None of the students had watched a peer friend prior to their own test drive.

2.4.1. Simulator test scenarios

To facilitate a comparison with the internship, the simulator test was developed in close collaboration with two train-driver instructors (one of which is the author of this paper) to mimic Swedish train driving reality as much as possible. Another aim was to develop a decisive and reliable test with many measuring points (i.e., test items). Hence, the test included both normal driving (approximately 30%–40% of the total time) and the handling of realistic, important, and relatively common special cases of different complexity, which should capture the eleven weeks of internship in a very comprised way. In sections of normal driving, the drivers manoeuvred the train according to lineside signals and in-cab signalling (e.g., catenary handling, honking before some level crossings, speed manoeuvring, and stops). Fortunately, actual train driving does not contain four special cases every 45 min – but it is reasonable for a simulator test to include it. The motive for this is partly to get more measuring points but also to facilitate a comparison with the supervisor grades as they are, to a great extent, based on the handling of special cases (directly linked to assessment areas 2, 3, 4, 6, 7, 11 and 12 used by the supervisors to assess the students, described in detail in section 2.5.2).

Each scenario also contained two special cases, one of which is simpler and one of which is more complex. The two more complex cases included more information from different sources to be handled by the driver and distractions from in-cab or outside information, which made an accurate situation analysis more difficult. The two simpler special cases did not include any other distraction. Since actual train driving contains both simpler and more complex special cases, it is essential that the simulator test also does so. Below is a description of the four special cases.

Scenario 1. Freight-train scenario

1. Simpler: A warning signal prior to a level-crossing signal gives the information that the level-crossing is not working correctly. This means that the level-crossing equipment does not for certain protect the road users.
2. Complex: The signal is red (meaning stop) because the points' position after the signal cannot be confirmed by the signal. This means that the signal needs to be passed after permission from the signaller and the point manually controlled. After the signal is passed, a man is standing near the track and close to the point. Finally, when driving at reduced speed, a minor balise transmission failure occurs (meaning that the train protection alarms but is still working properly) just before a lineside board informing about a catenary neutral section.

Scenario 2. Passenger train scenario

1. Simpler: Major balise transmission failure occurs at a signal inside an operation zone, meaning that the train protection system alarms and becomes temporarily inactive.
2. Complex: A main signal changes from red to green, but when the train passes, the information to the train from the balise is that the signal is red, leading to the train emergency brakes and the train protection system alarms. As a result, the train stops in a non-stopping area where regulations state that the pantograph must be lowered quickly to avoid a damaged catenary.

See the Appendix for a detailed list of events.

2.5. Performance measurement methods

When measuring overall driving performance, three methods have primarily been used in previous validation studies: a subjective measurement by an expert evaluator; objective protocols set to quantify driving errors; and performance data from programs connected to the simulator (Maag & Schmitz, 2012; Wynne et al., 2019). The two former were used in this study, and as in previous studies, LFS performance was operationalised as the number of driving errors (Tichon & Wallis, 2010; Korteling, Helsdingen, & Sluimer, 2016; Meuleners & Fraser, 2015; Mayhew et al., 2011; Selander, Stave, & Willstrand, 2019), while expert evaluators evaluated real train driving (Faschina et al., 2021; Galski et al., 1992; Aksan et al., 2016; Lee et al., 2003).

2.5.1. Simulator test performance

The same experts involved in the scenario development also created the driving errors protocol. As in previous studies, the objective was to create a protocol with expected actions and possible errors for each cue or incident (Tichon, 2007; Olsson et al., 2022). In total, the protocol consisted of 61 possible errors, of which 25 were rule-based (in comparison with the rules and regulations), 14 were skill-based (handling of the vehicle, the speed manoeuvring, and the train protection system), 16 were ineffective handling errors (time-costly train driving), and six were combinations of the former categories. A rule-based error (RBE) was, for example, noted when exceeding the speed limit in a situation where the speed should be reduced, if the driver did not call the signaller or gave the wrong information according to the rules. If an emergency brake occurred (intervened by the train protection system or by the driver) or the driver handled the train or train protection system incorrectly, a skill-based error (SBE) was detected. Ineffective performance, such as driving too slowly (more than 10 km/h below the speed limit), approaching a stop signal too slowly, or standing still unnecessarily, was registered as an ineffective handling error (IE). Each error had the same weight regardless of type or severity. Five of the possible errors in each scenario were not linked to any specific event but were of a more general nature, namely, driving too slowly, standing still without cause, intervention from the train protection system, an emergency brake intervention by the driver, and no deceleration check. To prevent these errors from being too decisive, a maximum of two errors were counted per scenario for the first four scenarios. A detailed description of the scenario cues, expected actions, and possible errors can be found in the Appendix.

2.5.2. Internship performance

The novice's internship was graded by experienced train drivers who were trained supervisors fulfilling the requirement of at least three years of train driver experience (Transportstyrelsen, 2011). The supervisors, blinded to the simulator test performance, used a standard checklist to facilitate the assessment, which has been used for grading train-driver students for many years. Each supervisor who supervised the student for at least five days during the internship period was asked to grade the student according to the checklist. Sometimes the student had several supervisors for at least five days, leading to several grades during one internship period.

Each of the twelve following areas was assessed: 1. *attitude and safety thinking in general*; 2. *retrieving and managing safety orders*; 3. *being on time and being aware of orders and timetables*; 4. *management of the train protection system and the route book*; 5. *handling the brakes, including deceleration checks*; 6. *attention to signals and train protection systems*; 7. *management of security calls*; 8. *driving techniques and attention to the train*; 9. *actions when the train stops at stations*; 10. *vehicle handling and signalling when shifting*; 11. *handling balise failures (problem with the train protection system)*; and 12. *handling of permitted passages of red signals and other special cases*.

To help the supervisors with the assessment, a more detailed description of the headline was included. For example, *attention to signals and train protection systems* means that *the student has complete control over signals, understands the information on ATC, and acts* *such that ATC does not have to intervene, while management of security calls* means that *the student conducts safety calls independently in a correct manner.*

The range of grades was 1 (lowest) to 5 (highest). In general, the criterion used for each area ranged from the student *not managing the task* (1) to the student *managing the task independently* (5), whereas 2–4 indicated that some help from the supervisor was required to manage the task.

If an area was impossible to grade because a specific situation had not occurred, no grade was submitted.

2.6. Analysis of data

Pearson product-moment correlations were used to assess the relative behavioural validity between performance in a simulator test and internship, internship time, and type of errors committed. Mean differences between classes were analysed with *t* tests for independent samples, with Cohen's *d* for effect size. Class × Error type was analysed with a 2 × 3 split-plot factorial ANOVA and post hoc independent samples *t* tests with Bonferroni adjustments for comparisons between groups per error type and Cohen's *d* for effect size. Main and interaction effects for class × peer watching were analysed with a 2 × 2 between-group ANOVA, with partial eta squared (η_p^2) as the effect size. The significance level was 5% for all analyses.

3. Results

In this section, the results are presented for each RQ respectively. Heading 3.1 answers RQ1 and RQ2, heading 3.2 presents results for RQ3, 3.3 for RQ4 and finally, 3.4 for RQ5.

3.1. Correlations between performance indicators, internship time, types of errors, and age

Thirty-four train driver students from two classes conducted the simulator test. One student from Class A was missing complete internship grades, and one student in Class B did not report the internship time. The internship grades are the average scores from internship periods 3–5 (11 weeks). Several supervisors sometimes graded the student in the same internship period; hence, more than one grade for each area could be noted. Other areas had not been graded by a supervisor in some internship periods since the basis for the assessment was too poor. The train driver students received between 25 and 55 grades for the three internships combined (M = 33.7, SD = 10.1). None of the students declared that they experienced simulator sickness during the test.

The number of driving errors in the simulator was negatively correlated with the internship grade (i.e., fewer errors usually meant a higher internship grade from the supervisors), r = -0.45, p <.05, and with the three error types (see Table 1). The internship grade was correlated with RBE and IE, while internship time and age failed to reach significance with any other variable.

3.2. Simulator test data

3.2.1. Descriptive statistics

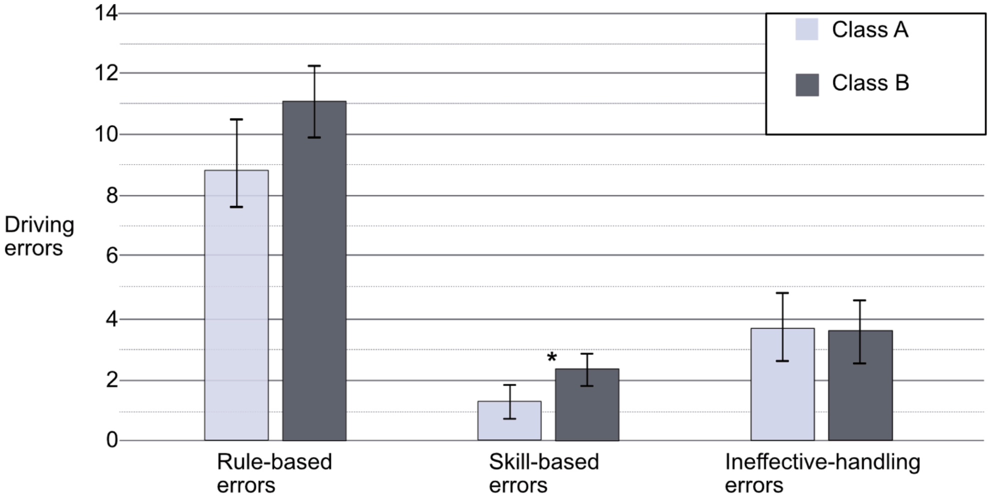
The 34 train driver students' average number of driving errors by class and error type are displayed in Fig. 3. When an error was classified as a combination of two error types, 0.5 errors per type were scored. For both classes combined, an average of 15.26 (SD = 4.65) errors out of 62 (25%) were committed. On average, the classes combined made 9.93 (SD = 3.02) rule-based errors (min = 2, one individual from Class A; max = 15.5, one individual from Class B), corresponding to 35 percent of the maximum. However, the drivers were only noted for 11 percent of the maximum possible skill-based errors (M = 1.77, SD = 1.18) and 21 percent of the ineffective handling errors (M = 3.57, SD = 2.14). The range for SBE was 0–4.5 (two drivers from Class A scored 0, while one from each class scored 4.5 SBEs). The range for IE was 0–8.5 (one student in Class B scored 0, while one from Class A scored 8.5 IEs).

Class B made 16.88 driving errors on average (SD = 3.64), which was 23 percent more than Class A (M = 13.65, SD = 5.09). Class B made 11.06 RBE on average (SD = 2.28), which was 25 percent more than that of Class A (M = 8.79, SD = 3.31). The largest difference between the classes (84 percent) concerned SBE, where Class A had M = 1.24 (SD = 1.11), and Class B had M = 2.29 (SD = 1.03) errors. For IE, almost no difference was observed between classes.

Table 1
Correlations between performance indicators, internship time, types of errors and age.

	1	2	3	4	5	6	7
1. Driving errors		**−0.45***	−0.09	**0.85***	**0.66***	**0.60***	0.15
2. Internship grade			0.34	**−0.35***	−0.10	**−0.43***	−0.15
3. Internship time				−0.09	−0.17	0.02	0.09
4. Rule-based errors					**0.51***	0.16	0.24
5. Skill-based errors						0.15	−0.10
6. Ineffective-handling errors							0.04
7. Age							

*p <.05.



*p < .05

Fig. 3. Mean number of driving error by error type and class, with 95-percent confidence intervals. *p <.05.

3.2.2. Inferential statistics

There was a main effect of Class on number of driving errors, $F(1, 32) = 4.55$, $MSE = 6.52$, $p = .04$, $\eta_p^2 = 0.12$, but no interaction effect, $F(1, 64) = 3.18$, $MSE = 4.41$, *ns*. There was a statistically significant difference ($p < .017$) between the classes concerning the number of SBEs, $t(32) = 2.89$, $p < .01$, $d = 0.99$, but not for RBEs, $t(32) = 2.33$, *ns.* nor IEs, $t(32) = 0.12$, *ns*.

3.3. Internship data

There were statistically significantly fewer hours of practice (12% less) and lower scores (6% lower) from the supervisor for Class B than for Class A (see Table 2). The lowest and highest numbers of driving errors for Class A were 4 and 22, respectively, while the driving errors for Class B ranged between 11 and 25, respectively. The range for Class A was 4.04–5. The only driver with an internship score below 4 (3.41) was from Class B, and this driver also had the third fewest internship hours and the four-highest number of driving errors. The highest grade within Class B was 4.87. Notably, all four students with less than 400 internship hours belonged to Class B. The age structure of the two groups was almost identical.

3.4. Effect of class and peer watching

There was a main effect of class on driving errors, $F(1, 30) = 4.34$, $MSE = 20.55$, $p < .05$, $\eta_p^2 = 0.13$. However, there was no main effect for peer watching, $F(1, 30) = 0.01$, $MSE = 20.55$, *ns.*, or an interaction effect for peer watching × class $F(1, 30) = 0.44$, $MSE = 20.55$, *ns*.

4. Discussion

Also in this section, the RQs are discussed separately, where the correlations (RQ1 and RQ2) are discussed under heading 4.1. 4.2 presents a discussion about RQ3, 4.3 is for RQ4 and under 4.4 RQ5 is discussed.

Table 2
Comparison between Classes A and B and descriptive and inference statistics.

	Class						
	A (n = 17)		B (n = 17)				
Variable	M	SD	M	SD	t	p	d
Internship grade	4.68*	0.28	4.40	0.33	2.66	0.01	0.93
Internship time	498.8	64.9	441.8*	87.4	2.14	0.04	0.74
Age	32.2	8.7	33.2	9.2	0.33	ns.	

*n = 16.

4.1. Ecological validity: Simulator test performance correlates with internship performance

The main contribution of this study is that the results suggest ecological validity between performance in a simulator test and 11 weeks of train driving in reality, which can be claimed to be a validation of the simulator test. No validation study of a simulator test for train driving was previously found, and just a few driving-simulator studies can claim to support ecological validity (with some exceptions, see, e.g., Hoffman & McDowd, 2010; Lew et al., 2005; Edwards, Hahn, & Fleishman, 1977). Clearly, and as stated by previous research, the differences between what can be measured in a low-fidelity simulator and the 11 weeks of train-driving it is trying to emulate are important to bear in mind (Allen, Mitchell, Stein, & Hogue, 1991). For example, fatigue or monotonous driving cannot be measured during a 45-minute test, nor can the to-be-driver's long-term attentiveness, carelessness, or stress tolerance. Hence, the medium-strong correlation should be considered a good result, although some similar validation studies of low- or mid-fidelity driving-simulators aiming for relative validity showed an even stronger correlation (e.g., Lee et al., 2003; Galski et al., 1992; as demonstrated in the review of automotive simulators by Wynne et al., 2019).

Likely, the high level of realism was crucial for the findings. As previous research recommends, realistic scenarios are sufficient and sometimes argued to be even more important than the level of physical fidelity to produce results transferable to actual driving (Meuleners & Fraser, 2015; Aronsson, Artman, Brynielsson, Lindquist, & Ramberg, 2021; Sellberg, Lindmark, & Rystedt, 2018). The realistic scenarios, together with the simulator's high functional fidelity (e.g., train dynamics in accordance with real trains, real levers, a replica of the real tracks, a highly functional train protection system, etc.) contributed to the same knowledge and skills being needed to perform the tasks in the simulator and during the internship. Thus, the results show that driver students who perform well in real-life driving also perform well in this simulator settings.

LFSs with a single screen showing the outside view are probably an even better fit for train driving (than, e.g., automotive, marine, or flight simulators) as a train follows the rail and does not need to be steered, either vertically or laterally, a factor where the field of view is considered important (Mecheri & Lobjois, 2018; Allen, Park, & Cook, 2010; Lidestam, Eriksson, & Eriksson, 2019). Likely, this explains the results from Tichon and Wallis (2010), where train drivers who had practised in a simulator with a single screen in front performed on par with drivers who had practised with a larger field of view in a simulator test.

Shechtman, Classen, Awadzi, and Mann (2009) argued that a validation study should measure multiple driving tasks, instead of just a small number of specific tasks, in order for simulator driving to be more consistent with the complexity of real driving. The results of this paper support that argument and show that it also applies to train driving. Hence, the present paper corroborates the results of previous driving simulator experiments that measured performance using multiple indicators. For example, the effect of experience has been studied in a car simulator using a large set of performance indicators, showing that experienced drivers made significantly fewer errors than novice drivers (Mayhew et al., 2011). Another experiment, also counting a wide variety of driving errors, showed that train drivers who practiced the new European signal and train protection system ERTMS in a simulator environment performed better (i.e., made fewer driving errors) than those who practiced in reality according to standard training methods (Olsson et al., 2022). The results from this study increase the validity of that study also for real train driving since a similar design for the simulator test and performance measurement was used.

4.2. Rule-based errors were the most common

The students generally committed few skill-based errors and considerably more rule-based errors. A probable conclusion is that the Swedish train driver education produces efficient students (IE) who are better at handling the train protection system, the vehicle, and manoeuvring the speed (SBE) than at concretising the regulations in a specific context (or lack of familiarity with regulations, RBE).

Class B made statistically significantly more SBEs than Class A, while the difference between RBEs and IEs was not significant (although RBEs were close). Class A's greater number of ineffective handling errors suggests that it is more a matter of attitude than knowledge that affects how highly the students value efficient driving.

4.3. COVID-19 restrictions negatively affected performance

Not surprisingly, the COVID-19 restrictions negatively affected performance, as the students from Class B made statistically significantly more driving errors and obtained a lower internship grade than those from Class A. As the result also showed a statistically significant difference regarding internship length, Class B's poorer performance is likely, in part, a result of the shortened internship. Nevertheless, a good guess is that the 10 weeks (out of 33) of distance education also affected the result to some extent. This conclusion is partly based on the fact that distance education likely has a negative effect on learning (depending on the extent of distance learning and the teacher's ability to maintain teaching quality, see Skolinspektionen, 2021) and that other conditions during the education were similar (e.g., relevant selection tests, syllabus, classroom environment, and teachers).

4.4. No effect of peer watching

That no statistically significant effect of peer watching was measured shows that, overall, the students' performance was probably unaffected. This should be seen as an encouraging result, as it shows that a peer-feedback educational method likely can be used without adversely affecting the driver's focus.

5. Conclusions

The main conclusion of this study is that a (physically) low-fidelity train-driving simulator with high levels of functional fidelity seems to be well suited for predicting real-life train-driving performance. LFS are often cheap and transportable (Roberts et al., 2020), which facilitates both research and practical usability, as more simulators can be purchased; furthermore, the simulators can travel to the users instead of the opposite. Therefore, the finding should have significant implications for both research and practical usability, and ultimately improve railway accident statistics. For research, this method with a standardized set of special cases could be used to study the effect of, for instance, fatigue or stress.

Additionally, TOCs may use the method to make visible and thereby remedy the knowledge deficiencies of train drivers. Detecting and being able to remedy knowledge gaps among the drivers at an early stage will decrease train delays and lower the risk of an actual accident. The present paper argues that this type of overall performance measurement has a better practical value than measuring only one specific performance indicator, such as speed manoeuvring or reaction time (which could be of equal importance) because it adds an opportunity for TOCs and researchers to evaluate drivers' overall practical train-driving skills in a safe and standardised way.

6. Limitations and suggestions for future studies

There are some limitations to be considered when interpreting the results. Firstly, a larger sample size would have increased the power to detect potentially important between-group effects this study could not find (e.g., effect of peer-watching on performance or effect of class on number of IEs or SBEs). Similarly, with a larger sample size, it would be possible to test the effect of different covariates (internship duration, age, etc.). Despite the heterogeneity of the participants concerning, for example, different ages and length of internship, the correlation between scores from real-life and simulator train-driving was moderately strong and statistically significant. Future studies might be able to control those other variables better and, thus, give a more certain answer to their effect on performance.

Unlike other driving professions, train driving is nationally specific, which may affect the generalisability of the results to other countries. The present study concerned Swedish regulations, signal and train protection systems, and language. It should, however, be possible to create the right circumstances in an LFS to adapt the methods from the present study to apply to other countries.

Another limitation concerns the study design regarding the effects of Covid-19 restrictions, i.e., distance education and a shortened internship time. An experiment with more isolated independent variables would increase the possibility of concluding the effects of a digitized theory and a shortened internship.

The results suggest that a poorer result in the simulator test leads to a greater risk of future accident since the simulator test correlated with real train-driving performance, and poorer performance increases the risk of accidents. However, to prove the long-time external validity of the test, it would be recommended to follow up the present study five years later with an accident report from the participating drivers (cf. Hoffman & McDowd, 2010).

Another suggestion for a future study is to test whether even higher correlations between scores from the simulator test and internship scores, respectively, would be observed if; a high-fidelity simulator (e.g., a full-cab simulator with a larger field of view and a moving base) was used or if; the same measurement method could be used for both the simulator test and internship or if; an extended simulator test was conducted (to also capture driver skills such as long-term attentiveness).

A continuation of this study would be to examine the effect of experience in a simulator test. Since the experience mainly consists of practical training, such a study would say something about how novices' training should be designed and how the length of the internship affects train-driving skills.

7. Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to confidentiality reasons but are available from the corresponding author on reasonable request.

Funding

This study was partly funded by the Swedish Transport Administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This research was funded by the Swedish Transport Administration. The author also wish to thank the educational company and the participants for their collaboration and contribution.

Author statement

PhD student Niklas Olsson is the sole author of this paper.

Appendix

Incident and cues	Expected action	Possible errors	Type of error
	Call signaller to receive permission to start (otherwise the signaller calls when the driver accelerates)	Do not call signaller prior to start	RBE
Receive permission to start	State the correct position and repeat the given permission	Do not repeat the permission given	RBE
Have received permission to start	Drive with reduced speed, max 35 km/h, until green signal is visible	Drive too fast	RBE
		Drive too slowly, less than 20 km/h	IE
Special case 1, simpler			
Warning signal prior to a level-crossing signal gives the information that the level-crossing is not clear.	Start braking before reaching the warning signal	Does not brake the train before reaching the signal	RBE
	Honk "train coming" initially	Do not honk initially	RBE
	Honk "train coming" repeatedly until the level-crossing-signal is visible or the train is certain to stop before level-crossing	Do not hor repeatedly	RBE
The level-crossing signal gives the information that the level-crossing is clear (level-crossing equipment works)	Release the brakes when the level-crossing signal is visible	Continue to brake the train even though the level-crossing signal indicates the level-crossing is clear	IE
	Call the signaller and inform about the incorrect warning signal and the correct level-crossing signal	Do not call the signaller about the incorrect warning signal	RBE
		Gives the wrong information about the level-crossing (i.e., place or what was incorrect)	RBE/ SBE
Driver receives information from lineside signal and in-cab signalling that the next main signal is red	Perform an effective deceleration towards the stop signal	Drive too slowly (less than 30 km/h at a certain point well before the signal)	IE
		Drive so fast that the train protection systems intervene	SBE
Special case 2, complex			
The stop signal is at stop because the point following the signal is out of control.	Call the signaller and receive a permission to pass the signal. The permission is received at a special form according to ceratin rules. The signaller calls the driver after 30 seconds.	Do not call the signaller for 30 seconds.	IE
		Do not perform a correct safety call and special form writing	RBE
The point supposed to be controlled ahead is in the correct position	Drive with reduced speed, max 35 km/h until the train reaches a green signal	Drive too fast	RBE
	Stop before the point	Do not stop before the point	RBE
	Drive with reduced speed over the point, max 10 km/h	Drive too fast over the point	RBE
A man is standing near the track close to the point.	Call the signaller and inform about the man standing near the track	Do not see/do not inform the signaller about the man	RBE/ SBE
A minor balise transmission failure occurs at the next signal right before a lineside sign that informs that the train is about to enter a catenary neutral section	Drive through the neutral section without stopping before the section	Stops the train before the neutral section (without the possibility of accelerating through the neutral section afterwards)	SBE
	Disconnect the train main power switch before entering the neutral section	Do not disconnect the train main power switch before entering the neutral section	RBE/ SBE
	Drive the train through the neutral section without connecting the trains main power switch	Do not stop the train while in the neutral section	RBE/ SBE
Possible errors not linked to certain events			
Drive according to the speed limits in sections of normal driving. Maximum two possible errors	Keep the speed in relation to speed limits	Drive too slowly, more than 10 km/h below speed limit	IE

(*continued on next page*)

(*continued*)

Incident and cues	Expected action	Possible errors	Type of error
will be counted (one per normal driving section)			
Drive the train according to the time-table (maximum two errors will be counted)	Do not stop unnecessary (for example to call the signaller or think about what to do)	Unnecessary standing still in another situation than presented above	IE
	Perform a deceleration check during the scenario	Do not perform deceleration check during the scenario	RBE
Drive too fast (maximum two errors will be counted)		Brake intervention from the train protection system	SBE
Drive too fast (maximum two errors will be counted)		The driver performs an unnecessary emergency brake	SBE

Scenario 2 - Passenger train scenario			
Incident and cues	**Expected action**	**Possible errors**	**Type of error**
The train's departure station, can see green signal	Enter the correct values in the train protection system	Entering incorrect values	RBE/ SBE
Special case 1, simpler			
Major balise transmission failure occurring at a signal inside an operation zone and the information from the train protection system ceases	Reduce the speed to maximum 40 km/h initially	Do not reduce the speed to below 40 km/h	RBE
	Drive a maximum of 45 km/h until new information occurs in the train protection system	Exceeds 45 km/h	RBE
	Drive a minimum of 30 km/h until new information occurs in the train protection system	Drive slower than 30 km/h	IE
New information occurs in the train protection system but is partial (info missing)	Drive a maximum of 80 km/h until new information occurs in the train protection system	Drive slower than 70 km/h	IE/RBE
	Produce the error code using the train protection system	Do not produce the error code	SBE
	Call the signaller about the balise failure	Do not call the signaller	RBE
	Give correct information about the balise error which includes error code, place and the type of error	Do not give the right information to the signaller	RBE
A lineside sign informs about a current catenary limitation section	Reduce the traction to zero or almost zero	Driver do not reduce the traction	RBE
Driver receives information from outside signal and in-cab signaling that the next main signal is red	Perform an effective deceleration against the stop signal	Drive too slowly (less than 30 km/h at a certain point well before the signal)	IE
		Drive so fast that the train protection systems intervene	SBE
a lineside sign informs that the train is about to enter a neutral section	Disconnect the trains power switch before entering the neutral section	Do not disconnect the train main breaker before entering the neutral section	RBE
Driver receives information from a lineside signal and in-cab signaling that the next main signal is red	Perform an effective deceleration against the stop signal	Drive too slowly (less than 30 km/h at a certain point well before the signal)	IE
		Drive so fast that the train protection system intervenes	SBE
Special case 2, complex			
A main signal turns from red to green but when the train passes the signal the balise gives the information stop leading to that the train emergency brakes. The train stops in a non-stopping area where regulations state that the pantograph must be lowered	Lower the pantograph within 7 seconds (otherwise there will be a loud bang illustrating a damaged catenary)	Do not lower pantograph	RBE/ SBE
	Call the signaller about the balise error within 30s	Do not call the signaller within 30s	RBE
	Ask the signaller for a special form according to the rules when this type of balise error occurs (otherwise the signaller will call the driver back about the special form)	Do not ask for the special form	RBE
	Perform a correct safety call and special form writing	Do not perform a correct safety call and special form writing	RBE
	Raise the pantograph and connect the train power switch after permission from signaller	Raises the pantograph and connect the main switch before given permission	RBE
	Release the brake within 10 seconds when trying to accelerate (otherwise the instructor will tell driver to do so)	Do not release the brake within 10 seconds when trying to accelerate	SBE

(*continued on next page*)

(*continued*)

Scenario 2 - Passenger train scenario			
Has received a permission to resume driving to the next signal from the signaller	Drive with reduced speed (maximum 35 km/h)	Exceeds 35 km/h	RBE
	Drive with a minimum of 20 km/h	Drive slower than 20 km/h	IE
Possible errors not linked to certain events			
Drive according to the speed limit in sections of normal driving. Maximum two possible errors will be counted (one per normal driving section)	Keep the speed in relation to speed limits	Drive too slowly, more than 10 km/h below speed limit	IE
Drive the train according to the time-table (maximum two errors will be counted)	Do not stop unnecessary (for example to call the signaller or think about what to do)	Unnecessary standing still in another situation than presented above	IE
	Perform a deceleration check during the scenario	Do not perform deceleration check during the scenario	RBE
Drive too fast (maximum two errors will be counted)		Brake intervention from the train protection system	SBE
Drive too fast (maximum two errors will be counted)		The driver performs an unnecessary emergency brake	SBE

References

Aksan, N., Hacker, S., Sager, L., Dawson, J., Anderson, S., & Rizzo, M. (2016). Correspondence between simulator and on-road drive performance: Implications for assessment of driving safety. *Geriatrics, 1*(1), 8.

Allen, R., Mitchell, D., Stein, A., & Hogue, J. (1991). *Validation of real-time man-in-the-loop simulation. VTI Report 372A, Part 4.* Linköping: Swedish Road and Traffic Research Institute.

Allen, R., Park, G., & Cook, M. (2010). Simulator fidelity and validity in a transfer-of-training context. *Transportation Research Record: Journal of the Transportation Research Board, 2185*(1), 40–47.

Andersson, A., Lidström, M., Peters, B., Rosberg, T., & Thorslund, B. (2017). *Framtagning av loktågsmodell för VTI:s tågsimulator [Production of locomotive train model for VTI's train simulator]*. Linköping: The Swedish National Road and Transport Research Institute.

Aronsson, S., Artman, H., Brynielsson, J., Lindquist, S., & Ramberg, R. (2021). Design of simulator training: A comparative study of Swedish dynamic decision-making training facilities. *Cognition, Technology & Work, 23*(1), 117–130.

Baysari, M., Caponecchia, C., McIntosh, A., & Wilson, J. (2009). Classification of errors contributing to rail incidents and accidents: A comparison of two human error identification techniques. *Safety Science, 47*(7), 947–957.

Bedard, M., Parkkari, M., Weaver, B., Riendeau, J., & Dahlquist, M. (2010). assessment of driving performance using a simulator protocol: Validity and reproducibility. *American Journal of Occupational Therapy, 64*(2), 336–340.

Blaauw, G. (1982). driving experience and task demands in simulator and instrumented car: A validation study. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 24*(4), 473–486.

Blana, E. (1996). driving simulator validation studies: A literature review. *Working paper, Institute of Transport Studies, University of Leeds, Leeds, UK*.

Brandenburger, N., & Jipp, M. (2017). Effects of expertise for automatic train operations. *Cognition, Technology & Work, 19*(4), 699–709.

Brandenburger, N., Hörmann, H., Stelling, D., & Naumann, A. (2017). Tasks, skills, and competences of future high-speed train drivers. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 231*(10), 1115–1122.

Branton, P. (1979). Investigations into the skills of train-driving. *Ergonomics, 22*(2), 155–164.

Buksh, A., Sharples, S., Wilson, JR., Coplestone A., & Morrisroe, G. (2013). A comparative cognitive task analysis of the different forms of drivingin the UK rail system. In N. Dadashi., A. Scott., J.R. Wilson., & A. Mills, *Rail human factors: Supporting reliability, safety and cost reduction* (pp. 173–182). London: Taylor and Francis.

Dahlstrom, N., Dekker, S., van Winsen, R., & Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science, 10*(4), 305–314.

Dorn, L., & Barker, D. (2005). The effects of driver training on simulated driving performance. *Accident Analysis & Prevention, 37*(1), 63–69.

Dorrian, J., Roach, G. D., Fletcher, A., & Dawson, D. (2007). Simulated train driving: Fatigue, self-awareness and cognitive disengagement. *Applied Ergonomics, 38*(2), 155–166.

Du, Y., Zhi, J., & He, S. (2022). Impact of experience on visual behavior and driving performance of high-speed train drivers. *Scientific Reports, 12*, 5956.

Dunn, N., & Williamson, A. (2012). Driving monotonous routes in a train simulator: The effect of task demand on drving performance and subjective experience. *Ergonomics, 55*(9), 997–1008.

Edkins, G., & Pollock, C. (1997). The influence of sustained attention on railway accidents. *Accident Analysis and Prevention, 4*, 533–539.

Edwards, D., Hahn, C., & Fleishman, E. (1977). Evaluation of laboratory methods for the study of driver behavior: Relations between simulator and street performance. *Journal of Applied Psychology, 62*(5), 559–566.

Evans, A. (2011). Fatal train accidents on Europe's railways: 1980–2009. *Accident Analysis & Prevention, 43*(1), 391–401.

Fan, C., Peng, Y., Peng, S., Zhang, H., Wu, Y., & Kwong, S. (2022). Detection of Train Driver Fatigue and Distraction Based on Forehead EEG: A Time-Series Ensemble Learning Method. *IEEE Transactions on Intelligent Transportation Systems, 23*(8), 13559–13569.

Faschina, S., Stieglitz, R.-D., Muri, R., Strohbeck-Kühner, P., Graf, M., Mager, R., & Pflueger, M. (2021). Driving errors, estimated performance and individual characteristics under simulated and real road traffic conditions – A validation study. *Transportation Research Part F: Traffic Psychology and Behaviour, 82*, 221–237.

Forsberg, R. (2016). Conditions affecting safety on the Swedish railway: Train drivers experiences and perceptions. *Safety Science, 85*, 53–59.

Freund, B., Gravenstein, S., Ferris, R., & Shaheen, E. (2002). evaluating driving performance of cognitively impaired and healthy older adults: A pilot study comparing on-road testing and driving simulation. *Journal of the American Geriatrics Society, 50*(7), 1309–1310.

Galski, T., Bruno, R., & Ehle, H. (1992). Driving after cerebral damage: A model with implications for evaluation. *The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association., 46*(4), 324–332.

Godley, S., Triggs, T., & Fildes, B. (2002). Driving simulator validation for speed research. *Accident Analysis & Prevention, 34*(5), 589–600.

Hamilton, W., & Clarke, T. (2005). Driver performance modelling and its practical application to railway safety. *Applied Ergonomics, 36*(6), 661–670.

Hamstra, S., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine, 89*(3), 387–392.

Harms, L., & Fredén, S. (1996). *Human and automatic train control in Scandinavian ATC-applications a pilot study.* Linköping: Swedish National Road and Transport Research Institute.

Hays, R. (1980). *Simulator fidelity: A concept paper.* U.S. Army Research Institute for the Behavioral and Social Sciences.

Hickey, A., & Collins, M. (2017). Disinhibition and train driver performance. *Safety Science, 95*, 104–115.

Hoffman, L., & McDowd, J. (2010). Simulator driving performance predicts accident reports five years later. *Psychology and Aging, 25*(3), 741–745.

Jansson, A., Olsson, E.-L., & Kecklund, L. (2000). *Att köra tåg: Lokförarens arbete ur ett systemperspektiv.* [To drive a train: The train driver's work from a system perspective]. Uppsala: Enheten för informationsteknologi, Uppsala Universitet.

Kaptein, N., Theeuwes, J., & van der Horst, R. (1996). Driving simulator validity: Some considerations. *Transportation Research Record, 1550*, 30–36.

Kecklund, L., Ingre, M., Kecklund, G., Söderström, M., Lindberg, E., Jansson, A., Olsson, E., & Sandblad, B. (2001). The TRAIN project: Railway safety and the train driver information environment and work situation—A summary of the main results. *Signalling Safety 2001.*

Korteling, J., Helsdingen, A., & Sluimer, R. (2016). An empirical evaluation of transfer-of-training of two flight simulation games. *Simulation & Gaming, 48*, 1–28.

Large, D., Golightly, D., & Taylor, E. (2017). Train-driving simulator studies: Can novice drivers deliver the goods? *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 231*(10), 1186–1194.

Lee, H. C., Cameron, D., & Lee, A. H. (2003). Assessing the driving performance of older adult drivers: On-road versus simulated driving. *Accident Analysis & Prevention, 35*(5), 797–803.

Lew, H., Poole, J., Lee, E., Jaffe, D., Huang, H.-C., & Brodd, E. (2005). Predictive validity of driving-simulator assessments following traumatic brain injury: A preliminary study. *Brain Injury, 19*(3), 177–188.

Lidestam, B., Eriksson, L., & Eriksson, O. (2019). Speed perception affected by field of view: Energy-based versus rhythm-based processing. *Transportation Research Part F: Traffic Psychology and Behaviour, 65*, 227–241.

Lobjois, R., Faure, V., Désiré, L., & Benguigui, N. (2021). Behavioral and workload measures in real and simulated driving: Do they tell us the same thing about the validity of driving simulation? *Safety Science, 134*, Article 105046.

Luke, T., Brook-Carter, N., Parkes, A. M., Grimes, E., & Mills, A. (2006). An investigation of train driver visual strategies. *Cognition, Technology and Work, 8*(1), 15–29.

Maag, C., & Schmitz, M. (2012). Assessment of train drivers' performance in a driving simulator. J.R, Wilson., A, Mills., T, Clarke., J, Rajan., N, Dadashi. *Rail Human Factors Around the World: Impacts on and of People for Successful Rail Operations* (pp. 136–145). CRC Press.

Mayhew, D., Simpson, H., Wood, K., Lonero, L., Clinton, K., & Johnson, A. (2011). On-road and simulated driving: Concurrent and discriminant validation. *Journal of Safety Research, 42*(4), 267–275.

Mecheri, S., & Lobjois, R. (2018). Steering control in a low-cost driving simulator: A case for the role of virtual vehicle cab. *Human Factors, 60*(5), 719–734.

Meuleners, L., & Fraser, M. (2015). A validation study of driving errors using a driving simulator. *Transportation Research Part F: Traffic Psychology and Behaviour, 29*, 14–21.

Mullen, N., Charlton, J., Devlin, A., & Bedard, M. (2011). Simulator validity: Behaviours observed on the simulator and on the road. in D, Fisher., M, Rizzo., J, Caird., & J, Lee. *Handbook of Driving Simulation for Engineering, Medicine and Psychology (1st ed)* (pp. 1–18). CRC Press.

Murphy, P. (2001). The role of communications in accidents and incidents during rail possessions. *Engineering Psychology and Cognitive Ergonomics, 5*, 447–454.

Myers, P., Starr, A., & Mullins, K. (2018). Flight simulator fidelity, training transfer, and the role of instructors in optimizing learning. *International Journal of Aviation, Aeronautics, and Aerospace, 5*(1), 6.

Naweed, A. (2013). Psychological factors for driver distraction and inattention in the Australian and New Zealand rail industry. *Accident Analysis & Prevention, 60*, 193–204.

Naweed, A. (2014). Investigations into the skills of modern and traditional train driving. *Applied Ergonomics, 45*(3), 462–470.

Noble, C. (2002). The relationship between fidelity and learning in aviation training and assessment. *Journal of Air Transportation, 7*(3), 33–54.

O'Connell, P., Lawton, F., Mills, A., & Klockner, K. (2017). Improving signal passed at danger management in New Zealand rail operations: Combining stabilised approach procedures with risk-triggered commentary driving. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 231*(10), 1070–1077.

Olsson, N., Lidestam, B., & Thorslund, B. (2022). Effect of train-driving simulator training in ERTMS, an experimental study, *0*(0). *Transportation Research Record*.

Phillips, R. O., & Sagberg, F. (2014). What did you expect? CREAM analysis of hazardous incidents occurring on approach to rail signals. *Safety Science, 66*, 92–100.

Roberts, A., Stanton, N., Plant, K., Fay, D., & Pope, K. (2020). You say it is physical, I say it is functional; let us call the whole thing off! Simulation: An application divided by lack of common language. *Theoretical Issues in Ergonomics Science, 21*(5), 507–536.

Roth, E., & Multer, J. (2009). *Technology implications of a cognitive task analysis for locomotive engineers United States*. Federal Railroad Administration. Office of Research and Development.

Saus, E., Johnsen, B., & Eid, J. (2010). Perceived learning outcome: The relationship between experience, realism and situation awareness during simulator training. *International Maritime Health, 62*(4), 258–264.

Schmuckler, M. (2001). What is ecological validity? *A dimensional analysis. Infancy, 2*(4), 419–436.

Selander, H., Stave, C., Willstrand, T.D., Peters., B. (2019). Driving simulator-based training to improve self-rating ability of driving performance in older adults: A pilot study. *European Transport Research Review 11(1),* 35.

Sellberg, C., Lindmark, O., & Rystedt, H. (2018). Learning to navigate: The centrality of instructions and assessments for developing students' professional competencies in simulator-based training. *WMU Journal of Maritime Affairs, 17*(2), 249–265.

Shanahan, P., Gregory, D., Shannon, M., & Gibson, H. (2007). The role of communication errors in railway incident causation. In J. Wilson, B. Norris, T. Clarke, & A. Mills, *People and Rail Systems: Human Factors at the Heart of the Railway* (pp. 427–438). Aldershot, England: Ashgate.

Shechtman, O., Classen, S., Awadzi, K., & Mann, W. (2009). Comparison of driving errors between on-the-road and simulated driving assessment: A validation study. *Traffic Injury Prevention, 10*(4), 379–385.

Skolinspektionen. (2021*). Fjärr- och distansundervisningens konsekvenser. Granskning av hur fjärr- och distansundervisningen under covid-19-pandemin påverkat elevers förutsättningar att nå målen och deras hälsa.* [The consequences of distance learning. Review of how distance learning during the COVID-19 pandemic has affected students' ability to achieve the goals and their health] Stockholm: Skolinspektionen.

Smith, P., Kyriakidis, M., Majumdar, A., & Ochieng, W. (2013). Impact of european railway traffic management system on human performance in railway operations: European findings. *Transportation Research Record: Journal of the Transportation Research Board, 2374*(1), 83–92.

Suter, J., & Stoller, N. (2014). Use of simulators to investigate complex issues relating to human factors. *FORMS/FORMAT 2014*: *10th Symposium on Formal Methods for Automation and Safety in Railway and Automotive Systems* (pp. 86–96).

Tabai, B. H., Bagheri, M., Sadeghi-Firoozabadi, V., & Shahidi, V. (2017). The relationship between train drivers' attention and accident involvement. *4th International Conference on Transportation Information and Safety (ICTIS)* (pp. 1034–1039).

Thorslund, B., Rosberg, T., Lindström, A., & Peters, B. (2019). User-centered development of train driver simulator for education and training. *8th International Conference on Railway Operations Modelling and Analysis (ICROMA 2019).* Norrköping, Sweden.

Tichon, J. (2007). The use of expert knowledge in the development of simulations for train driver training. *Cognition, Technology and Work, 9*(4), 177–187.

Tichon, J., & Wallis, G. (2010). Stress training and simulator complexity: Why sometimes more is less. *Behaviour & Information Technology, 29*(5), 459–466.

Transportstyrelsen. (2011). *Transportstyrelsens föreskrifter om förarutbildning m.m. enligt lagen (2011:725) om behörighet för lokförare* [The Swedish Transport Agencýs regulations on driver training etc. in accordance with the Act (2011:725) on qualifications for train drivers]. Norrköping, Sweden: Transportstyrelsen. Retreived from: transportstyrelsen.se/tsfs/TSFS%202011_60.pdf.

van der Weide, R. (2017). ERTMS roll out from a train driver's perspective. *Sixth International Human Factors Rail Conference. London, UK* (pp. 6–9).

Verstappen, V. (2017). The performance of Dutch train drivers based on the impact of the presence of a second person in the cab. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 231*(10), 1130–1140.

Wynne, R., Beanland, V., & Salmon, P. (2019). Systematic review of driving simulator validation studies. *Safety Science, 117*, 138–151.

de Winter, J., de Groot, S., Mulder, M., Wieringa, P., Dankelman, J., & Mulder, J. (2009). Relationships between driving simulator performance and driving test results. *Ergonomics, 52*(2), 137–153.

Zhang, X., Li, J., Liu, Y., Zhang, Z., Wang, Z., Luo, D., … Wang, C. (2017). Design of a Fatigue Detection System for High-Speed Trains Based on Driver Vigilance Using a Wireless Wearable EEG. *Sensors, 17*(3), 486.