A hybrid FRAM-Kirkpatrick framework for evaluating OSH digital training effectiveness in complex socio-technical systems

Andrei Stefan Iacob[1], Oana-Roxana Chivu[1]*, Mihnea Costoiu[1], Alina Trifu[2]*, Roland Moraru[3], Adrian Bibiș[4]

[1]National University of Science and Technology POLITEHNICA Bucharest, Faculty of Industrial Engineering and Robotics, 313 Splaiul Independenței, Sector 6, Zip Code 060042, Bucharest, Romania

[2]National Research and Development Institute on Occupational Safety - I.N.C.D.P.M. "Alexandru Darabont", 35A Ghencea Blvd., Zip Code 061692, Bucharest, Romania

[3]University of Petroșani, Faculty of Mining, Department of Management and Industrial Engineering, Petrosani, Zip Code 332006, Hunedoara, Romania

[4]National Research and Development Institute for Gas Turbines COMOTI, B-dul Iuliu Maniu nr. 220D, sector 6, Zip Code 061126, Bucharest, Romania

Corresponding authors

Received: December 3, 2025. Revised: April 4, 2026. Accepted: May 9, 2026. Published: June 30, 2026.

Abstract - Digital occupational safety and health training is nowadays widely adopted in complex socio-technical systems. Yet its effectiveness is often assessed using linear approaches that fail to capture impacts on real work and system performance. This study argues that training effectiveness cannot be meaningfully evaluated without a systemic perspective. To address this gap, the paper proposes the FRAM-Kirkpatrick Framework, integrating the Functional Resonance Analysis Method (FRAM) with the Kirkpatrick training evaluation model. The framework conceptualizes digital OSH training as a systemic intervention influencing functional performance variability. The FRAM-Kirkpatrick Framework is applied through multiple case studies in the road transport sector involving three companies and 47 professional drivers. Results show significant improvements in learning and safe behavior, accompanied by a substantial reduction in functional variability within critical system functions. The findings demonstrate that the FRAM-Kirkpatrick Framework enables a more robust evaluation of digital training effectiveness in complex socio-technical systems.

Keywords - Complex socio-technical systems, Digital Training, Occupational Safety and Health, Functional Resonance Analysis Method (FRAM), Kirkpatrick training evaluation model, road transport.

I. INTRODUCTION

COMPLEX socio-technical systems are characterized by interactions between human operators, technologies, organizational structures, and regulatory constraints. In such systems, safety and performance result from daily adaptations rather than strict adherence to procedures, [1], [2]. Road transport exemplifies this complexity: drivers operate under time pressure, variable traffic conditions, fatigue, and regulatory requirements, while organizations must ensure continuous compliance with occupational health and safety (OHS) requirements, [3], [4].

Digital training has been increasingly promoted as a solution for improving safety performance, flexibility, and compliance, [5], [6]. However, in many organisations, digitalisation is formally replacing traditional classroom training without fundamentally changing how training effectiveness is understood or assessed, [7]. Conventional assessment approaches, often limited to participant satisfaction or short-term knowledge tests, do not adequately reflect how training impacts actual work practices, functional linkages, or systemic safety outcomes, [8], [9].

This paper aims to demonstrate that the effectiveness of training in complex socio-technical systems cannot be meaningfully assessed without a systemic perspective, [10], [11]. To address this, we propose a hybrid framework, the Functional Resonance Analysis Model (FRAM) - Kirkpatrick Model for training evaluation, in which:

- FRAM is used to analyze system functions, variability, and resonance mechanisms, [12], [13];
- Kirkpatrick model provides a structured empirical validation of training effects, [14], [15];
- Digital training is treated as a systemic intervention capable of remodeling functional parameters and reducing unwanted variability, [16], [17].

II. State-of-the-art

A. Digital training in complex socio-technical systems

The digital transformation of training processes has become a central element in occupational safety and health (OSH) management, especially in areas characterized by high risk and operational complexity. Such areas are those in road transport, aviation, the energy industry, or construction. Digital training includes a wide range of technologies and approaches, from e-learning and micro-learning platforms to advanced simulations, virtual reality (VR), and adaptive systems based on operational data, [18], [19], [20].

A major advantage of digital training is its scalability and flexibility, allowing continuous training of a large number of workers, in different geographical contexts and with a reduced impact on operational activity, [21] [22]. In addition, digital technologies facilitate the personalization of content and the integration of real-time feedback. This is particularly relevant in socio-technical systems, where working conditions are dynamic and unpredictable, [23].

However, the literature highlights that the effectiveness of digital training is not guaranteed by the mere digitization of content, [24] [25]. Numerous studies show that the transfer of knowledge and behaviors to real work depends on contextual factors such as organizational culture, time pressures, technological constraints, and the level of autonomy of operators, [26] [27]. Without an alignment between training and the reality of work ("work-as-done"), digital training risks becoming a formal compliance tool, without a significant impact on safety, [28] [29].

In the context of complex socio-technical systems, training should be understood not only as an educational process but as a mechanism for regulating human adaptation. Operators are required to manage daily variability in their work, and training influences how they anticipate, compensate for, and respond to unexpected situations, [30]. Thus, the evaluation of training should go beyond individual performance and include the effects on the behavior of the system as a whole.

D. Functional Resonance Analysis Method (FRAM)

The Functional Resonance Analysis Method (FRAM) was developed to address the limitations of traditional accident and safety analysis approaches that focus on linear causal chains and individual errors. The method provides a systemic perspective on performance and safety, [31] [32]. FRAM proposes a non-linear view, in which the performance of a system results from the interaction of its functions and their inherent variability. This variability is considered a necessary feature for adapting to changing operational conditions, [33].

In FRAM, each function is characterized by six fundamental aspects (Input, Output, Preconditions, Resources, Control, and Time). Risks arise when the variability of several functions combines and amplifies, generating negative functional resonance, [34]. FRAM has been successfully applied in areas such as aviation, rail transport, the medical sector, and the energy industry, providing explanations for the occurrence of accidents and incidents, [35], [36].

The application of FRAM in the evaluation of training and organizational interventions is still limited, one of the main challenges being the difficulty of empirically validating the analyses. Although the method is extremely valuable for understanding systemic mechanisms, its results are often qualitative or semi-quantitative and based on expert opinions. This makes it difficult to demonstrate the direct effect of interventions, such as digital training, on the overall performance of the system, [37].

C. Kirkpatrick Model

The Kirkpatrick model is one of the most widely used frameworks for evaluating training effectiveness. It is used in areas such as human resource management, vocational education, and occupational safety and health (OSH) training, [38]. The model consists of four levels: Level 1 (L1) - participant response, Level 2 (L2) - learning, Level 3 (L3) - behavioral change, and Level 4 (L4) - organizational outcomes. The conceptual clarity and simplicity of the Kirkpatrick model have contributed significantly to its popularity in practice, providing a common language for evaluating training among specialists, managers, and decision-makers, [39].

The academic literature has highlighted a series of structural limitations of the model, especially when applied to complex socio-technical systems. A recurring criticism concerns the assumption of a linear and causal relationship between the four levels, although empirical evidence shows that the link between learning, behavior, and outcomes is often mediated by organizational and contextual factors that are not explicitly captured by the model, [40].

In addition, the ambiguity of level 4 (Outcomes) is frequently reported, as the indicators used at this level are influenced by numerous factors external to the training, making it difficult to directly attribute changes to the educational process alone. Consequently, the use of level 4 as a direct measure of training effectiveness is often contested, especially in contexts characterized by functional interdependencies and continuous adaptation.

Gap identified in the research. The literature review highlights a significant gap at the intersection of training evaluation and systemic analysis of safety in complex socio-technical systems. On the one hand, established training evaluation models, such as Kirkpatrick, provide measurable and easily applied indicators, but are limited in explaining the mechanisms by which training influences the behavior of the system as a whole. On the other hand, systemic methods such as FRAM provide a deep understanding of functional variability and resonance, but are rarely used for empirical evaluation of the effectiveness of organizational interventions, including digital training. The existing literature generally treats training as an activity separate from the functional structure of the system. It does not analyze how it can change the parameters of functions, the couplings between them, or the overall level of variability. There is a lack of methodological frameworks that allow for the explicit correlation of training evaluation results with observable changes in system behavior and risk reduction. This gap is particularly relevant in the context of digital training, where data is abundant but its systemic interpretation remains limited.

III. METHODOLOGY

The study adopts a mixed-methods design, combining FRAM functional analysis, quantitative training evaluation using the Kirkpatrick model, and comparative pre-intervention and post-intervention analysis. The methodological framework is applied in the context of digital occupational safety and health (OSH) training, in a socio-technical system characterized by high operational variability (road transport).

A. Conceptual architecture of the FRAM–Kirkpatrick hybrid framework for evaluating digital instruction

Starting from the identified research gap, this study proposes a hybrid conceptual framework that integrates the Functional Resonance Analysis Method (FRAM) with the Kirkpatrick model, in order to evaluate digital training as a systemic intervention in complex socio-technical systems. The architecture of the framework is built on the fundamental assumption that the effectiveness of training cannot be evaluated in isolation, at the individual level. It must be analyzed in relation to the changes induced in the functional structure of the system and in the variability of its performance.

The proposed framework is organized in an iterative cycle of evaluation and optimization, in which digital training, FRAM functional analysis, and Kirkpatrick empirical evaluation are interconnected in a bidirectional relationship.

B. Digital training as a systemic intervention

Within the proposed architecture, digital training is not treated as a simple educational process, but as a systemic intervention capable of influencing the parameters of socio-technical functions. It acts on:

- cognitive resources of operators (knowledge, skills, situational awareness);
- control mechanisms (compliance with procedures, decision-making);
- functional preconditions (mental preparation, motivation, availability for action);
- operational timing (anticipation and management of time pressures).

In this study, digital training is conceptualized as a factor regulating the variability of everyday performance ("performance variability") and not as a tool for rigid standardization.

C. The FRAM component of the framework: modeling functional structure and variability

The FRAM component of the hybrid framework is used to describe and analyze the functional structure of the socio-technical system. This is achieved by identifying relevant functions and the interdependencies between them. In the architecture of the proposed framework, FRAM fulfills three main roles:

- identification of critical functions on which training can have an impact;
- modeling of functional variability before and after the implementation of training;
- analysis of functional resonance mechanisms, which can lead to unsafe results.

The FRAM functions are identified through a complex analysis of operational processes, interviews with OSH experts, procedural documentation, and observations of real activity (work-as-done). Examples of functions are: “Safe driving”, “Fatigue management”, “Compliance with OSH procedures”.

Each function f is described by the set:

$$f=\{I,O,P,R,C,T\} \quad (1)$$

where: I = Input, O = Output, P = Preconditions, R = Resources, C = Control, T = Time

The total variability Vf of a FRAM function is calculated as:

$$Vf = \sum_{i=1}^{n} wi\,\Delta Pi \quad (2)$$

where: ΔPi = variation of functional parameter i, wi = weight of functional parameter i

This formula allows both the before/after training comparison and the analysis of the effect of training on functional stability.

D. Kirkpatrick component: empirical validation of training effects

The Kirkpatrick model is integrated into the hybrid framework to provide structured empirical validation of the effects of digital training. In the proposed architecture, the Kirkpatrick levels are used as follows:

L1 - Level 1 (Response) is interpreted as an indicator of the acceptance and quality of the FRAM Preconditions for the application of training. Indicator used: standardized questionnaires.

L2 - Level 2 (Learning) reflects the consolidation of cognitive resources, relevant to the Resources parameter of FRAM. Indicator used: knowledge test scores

L3 - Level 3 (Behavior) is directly associated with the stability of Functional Outputs and the reduction of behavioral deviations. Indicator used: observational operational indicators
L4 - Level 4 (Outcomes) is used as a systemic validation indicator and not as a direct measure of training effectiveness. Indicator used: safety indicators (incidents, near-misses).

E. Operational integration of the Kirkpatrick model with the FRAM functional structure / Correlation of Kirkpatrick levels with FRAM functional components

The integration of the two models is achieved by correlating the Kirkpatrick assessment levels with the structural elements of a FRAM function (Input, Output, Preconditions, Resources, Control, and Time). This mapping allows the association of training assessment indicators with sources of functional variability (Table 1).

Table 1. Correlation of Kirkpatrick levels with FRAM components

Kirkpatrick Level	*Evaluation objective*	*Dominant FRAM component*	*Associated indicators*
Level 1 - Reaction	Acceptance of training	Preconditions	Satisfaction index, perceived usefulness
Level 2 - Learning	Knowledge acquisition	Resources / Control	Knowledge retention, test scores
Level 3 - Behavior	Transfer into practice	Output	Behavior change
Level 4 - Results	Systemic impact	Functional Resonance	System performance

Source: created by the authors.

Through this correlation, training evaluation results can be used to justify changes in the FRAM functional structure and to demonstrate, in a quantifiable way, the reduction of unwanted variability. The integration of the Kirkpatrick model with the FRAM method is the central element of the proposed hybrid framework, as it transforms two conceptually different approaches into a coherent tool for analyzing and evaluating digital training in complex socio-technical systems. This integration does not imply a simple joining of the two models, but a systematic correlation between the levels of training evaluation and the sources of functional variability identified by FRAM.

Table 1 represents the key tool of operational integration, as it translates training assessment results into functional variability. Each row of the table describes a causal relationship between a Kirkpatrick assessment level and a dominant FRAM component, as well as the indicators used to measure it. It should be noted that the FRAM component Input is not explicitly mapped to any Kirkpatrick level within the framework. This is a deliberate methodological choice. In the proposed framework, digital training is conceptualized as an intervention acting on functional performance rather than on operational demands. Consequently, Input is treated as an exogenous variable, reflecting task requirements and system constraints that are not directly modified by training. Training influences how operators respond to inputs through changes in resources, control mechanisms, preconditions, and observable outputs, rather than altering the inputs themselves.

Level 1 Kirkpatrick (Reaction) ↔ Preconditions (FRAM). Level 1 assesses the participants' reaction to the training, including satisfaction, perceived usefulness, and perceived relevance to the real work. In the FRAM framework, these elements are correlated with Preconditions, as they reflect the psychological and cognitive readiness of operators to use the acquired skills. A negative reaction or a low perception of the usefulness of the training constitutes a deficient precondition, which can amplify the variability of other functions, even if the training is well designed from a technical point of view. Thus, the results of Level 1 can justify changes in the FRAM structure by:

- redefining the preconditions for critical functions;
- introducing additional organizational constraints (e.g., mandatory training, feedback sessions);
- adjusting the training design to reduce the variability induced by the lack of acceptance.

Kirkpatrick Level 2 – Learning ↔ Resources / Control (FRAM). Level 2 reflects the degree of knowledge acquisition and retention, being directly correlated with Resources and Control in FRAM. In this context, knowledge, skills, and competences acquired through training are treated as functional resources, and the ability to correctly apply procedures and rules is associated with the control mechanisms of the functions. Poor results at Level 2 indicate insufficient resources or ineffective control mechanisms, which can lead to: increased variability of execution time; inappropriate use of technology; improvised decisions under pressure. Therefore, data from Level 2 can be used to justify recalibrating the resources allocated to FRAM functions, introducing additional controls (digital guides, decision support systems), and redefining the acceptable limits of functional variability.

Kirkpatrick Level 3 – Behavior ↔ Output (FRAM). Level 3 is essential for operational integration, as it reflects the effective transfer of training into observable behaviors in real activity. In FRAM, this level is correlated with Output, since behavioral deviations represent direct manifestations of functional variability. The reduction of behavioral deviations observed after training indicates a stabilization of the outputs of critical functions, which reduces the probability of the appearance of unstable couplings between functions. In FRAM terms, this allows adjusting the coupling relationships between functions, eliminating redundant or unstable connections, and redefining the conditions for accepting outputs. Thus, Level 3 provides the most relevant data for justifying structural changes in the FRAM model.

Kirkpatrick Level 4 – Results ↔ Functional Resonance (FRAM). Level 4 reflects the impact of training on systemic performance, being correlated with the concept of functional resonance in FRAM. The indicators used (incidents, performance, safety) are not associated with a single function, but result from the interaction of several functions and the way in which their variability combines. The decrease in negative indicators at Level 4 constitutes a systemic validation of the changes made to the FRAM structure and confirms that the reduction of variability at the functional level has translated into a real improvement in performance and safety.

F. Mathematical formalization of the impact of training on functional variability

F.1. Reducing unwanted variability

By systematically correlating Kirkpatrick levels with FRAM components, the proposed framework allows for the quantification of the effect of training on functional variability. Kirkpatrick indicators become input variables for recalculating the variability of FRAM functions, and the observed changes can be expressed numerically by the global variability index V_f.

$$Vf = \sum_{i=1}^{n} wi\,\Delta Pi \qquad (3)$$

This expression provides an aggregate representation of functional variability, allowing comparison of the state of the system before and after the implementation of an intervention.

V_f – total variability of the FRAM function. It represents the total variability of the performance of a FRAM function, expressed as a scalar value. This value summarizes the contributions of all relevant functional parameters and can be used as an indicator of the stability or instability of the respective function. Higher values of Vf indicate an increased level of variability and, implicitly, a higher potential for unstable coupling with other functions in the system. The reduction of the Vf value after training constitutes a quantifiable demonstration of the effectiveness of the intervention, surpassing traditional assessments based solely on satisfaction or test scores. Thus, digital training is validated not only as an educational process but also as an intervention to control variability and prevent negative functional resonance.

ΔPi – variation of functional parameters. The term ΔPi describes the variation of a specific functional parameter associated with the analyzed function. Within the FRAM, these parameters correspond to fundamental aspects of a function, such as:

- *Time* – variations in execution time*;*
- *Resources* – fluctuations in the availability or quality of resources*;*
- *Control* – variations in monitoring mechanisms, rules, or procedures*;*
- *Preconditions* – changes in the initial conditions necessary for the activation of the function;
- *Input / Output* – inconsistencies or deviations in the inputs and outputs of the function.

The variation of ΔPi can be expressed by quantitative indicators (e.g., standard deviations, error rates, time delays) or by normalized scores obtained from operational assessments. Thus, the term reflects the amplitude of instability associated with each parameter.

wi – weight of the functional parameter. The coefficient wi represents the relative importance of parameter i in determining the total variability of the function. Not all parameters contribute equally to the functional behavior; for example, variations related to Control or Time may have a more pronounced impact on safety than minor variations of Resources. In the proposed framework, digital training is conceptualized as a systemic intervention that directly influences the values of ΔPi and, implicitly, the global value of V_f. Training does not eliminate variability, but shapes and constrains it within acceptable limits, improving the ability of operators to manage variable situations.

F.2 Training efficiency

To capture this effect, the training efficiency Et is defined, derived from the Kirkpatrick levels relevant to operational performance:

$$E_t = \alpha L2 + \beta L3 \qquad (4)$$

where: L2 reflects the level of learning (knowledge acquisition); L3 reflects behavioral transfer in real activity; α and β are weighting coefficients

F.3 The impact of training on functional variability

The effect of training on functional variability is expressed by the relationship:

$$V_f^{trained} = V_f^{intial} \cdot (1 - E_t) \qquad (5)$$

This formulation reflects the fact that more effective training leads to a proportional reduction in functional variability. Conceptually, the reduction in V_f indicates:

- stabilization of Control and Time parameters;
- reduction of behavioral deviations (Output);

increase in predictability of function under variable conditions.

IV. Case Study - Application of the FRAM–Kirkpatrick Hybrid Framework in Road Transport

A. Quantitative analysis design and working hypotheses

To validate the FRAM–Kirkpatrick hybrid framework, the quantitative analysis was built on the fundamental hypothesis that digital occupational safety and health (OSH) training influences the performance of the socio-technical system by reducing functional variability, and this reduction can be measured indirectly through the levels of the Kirkpatrick model. The following operational hypotheses were formulated:

- H1: Improvements at Kirkpatrick levels L2 (learning) and L3 (behavior) lead to an increase in training efficiency Et;
- H2: Increased training efficiency Et is associated with reduced FRAM functional variability Vf;
- H3: Reduced functional variability Vf is reflected in improvements in safety outcomes (Kirkpatrick L4).

In this study, training efficiency was operationalized as a composite indicator derived from Kirkpatrick Levels 2 (Learning) and 3 (Behavior). This choice is grounded in the training evaluation literature, which consistently identifies learning outcomes and behavioral transfer as the primary direct effects of training interventions, [34]. A weighted linear aggregation was employed to ensure transparency and interpretability, in line with established practices for constructing composite performance indicators.

It should be mentioned that Kirkpatrick Level 4 was not included in the calculation of training efficiency. This choice is based on the distinction between training effectiveness at the operational level and system-level outcomes. While Levels 2 and 3 reflect direct effects of training on knowledge acquisition and behavioral change, Level 4 represents a lagging organizational outcome influenced by multiple contextual factors. Consequently, Level 4 was used to validate the systemic impact of training through observed reductions in functional variability and safety events, rather than as a direct component of training efficiency.

B. Description of the socio-technical system analyzed

The case study focuses on a road transport system characterized by complex interactions between professional drivers, infrastructure, vehicles, digital support systems, and organizational structures. Road transport activity is carried out in a dynamic environment, marked by time pressures, operational variability, and high exposure to road safety risks.

From an occupational health and safety perspective, the main risks identified include: human errors caused by fatigue and stress; failure to comply with operational procedures; inappropriate use of digital assistance systems (ADAS, digital tachograph); high variability of traffic and environmental conditions. In this context, digital driver training is a critical intervention to reduce performance variability and increase the stability of the socio-technical system.

The quantitative analysis of the FRAM-Kirkpatrick Framework (FKF) was conducted through a multiple case study applied to three road freight transport companies in Central and Eastern Europe. The organizations were selected based on common criteria, including operating a fleet of over 50 heavy vehicles, carrying out international long-distance transport activities, implementing a digital training program in the field of occupational safety and health, and the availability of operational data relevant to safety. To comply with confidentiality requirements, the companies are presented anonymously as Company A (international general transport), Company B (refrigerated transport), and Company C (industrial materials transport). The analysis covered 12 months, structured into a 6-month pre-intervention phase, characterized by traditional or limited training, and a 6-month post-intervention phase, in which the digital training assessed by the FKF framework was implemented.

C. Description of the intervention: digital OSH training

The intervention analyzed consisted of implementing a digital training program in the field of occupational safety and health (OSH), addressed to professional drivers from the three transport companies included in the study. The program was designed as a structured intervention and integrated e-learning modules on occupational risks, specific training on fatigue management and compliance with driving times, as well as video simulations of critical situations encountered in real activity. In addition, the training included periodic automatic assessments and individualized feedback mechanisms, based on operational data collected from daily activity. The program was applied to a total sample of 47 drivers, distributed between Company A (18 drivers), Company B (14 drivers), and Company C (15 drivers). For quantitative analysis and to protect the confidentiality of the participants, individual data were aggregated into performance profiles P1-P5, which represent clusters of operational behavior.

D. Operationalizing FRAM in case studies

For each company, a FRAM model of safe driving activity was built, including functions such as: F1 - Trip planning; F2 - Vehicle driving; F3 - Fatigue management; F4 - Compliance with SSM procedures; F5 - Reaction to unforeseen situations. Each function was described by the six FRAM aspects (I, O, P, R, C, T), and the variability of the parameters was estimated by telematics data (timing, deviations), SSM reports, operational observations, and calibrated expert judgment.

E. Training Evaluation – Kirkpatrick L2 and L3 Data

Kirkpatrick Level 2 (L2) – Learning

Level 2 of the Kirkpatrick model was used to assess the acquisition and consolidation of knowledge gained from digital training. The L2 assessment was carried out through standardized digital tests, integrated into the training platform used by the participating companies. The scores obtained by the participants before and after the implementation of the training program were extracted from the learning management system and aggregated at the organizational level. Figure 1 and Table 2 summarize the mean pre- and post-training values, which were subsequently used as input variables in the quantitative analysis within the FRAM-Kirkpatrick Framework.

Table 2. L2 results (aggregated averages by company)

Company	L2 pre-training	L2 post-training	ΔL2
A	60.8	82.1	+21.3
B	63.2	84.5	+21.3
C	58.9	81.6	+22.7
Medie	61.0	82.7	+21.7

Source: created by the authors.

Fig.1. L2 mean pre-training and post-training values. *Source: created by the authors.*

Kirkpatrick Level 3 (L3) - Behavior

Level 3 of the Kirkpatrick model was used to assess the transfer of digital training into real operational behavior relevant to safety. The L3 assessment was based on objective operational indicators, extracted from telematics systems and internal occupational health and safety reports. These included non-compliance with rest periods, sudden braking, speeding, and procedural non-compliance. The values of these deviations were aggregated and normalized at the company level, expressing the average number of events per driver and per month. Figure 2 and Table 3 present the comparative pre- and post-training values used in the subsequent quantitative analysis within the FRAM-Kirkpatrick Framework.

Table 3. Behavioral deviations (averages/driver/month)

Company	*L3 pre-training*	*L3 post-training*	*ΔL3*
A	3.4	1.6	−1.8
B	3.1	1.4	−1.7
C	3.6	1.7	−1.9

Source: created by the authors.

Fig. 2. Comparative L3 pre- and post-training values. *Source: created by the authors.*

F. Calculating training efficiency

The effectiveness of training was calculated by defining a composite indicator Et, which integrates the results of the Kirkpatrick levels L2 (Learning) and L3 (Behavior), according to the relationship.

$$Et = 0.5 \cdot L2 + 0.5 \cdot L3 \quad (6)$$

Equal weights were used to reflect the balanced importance of knowledge acquisition and behavioral transfer in assessing the effectiveness of digital instruction. L3 scores were normalized before calculation, based on observed behavioral deviations, to ensure comparability with L2 values.

$$Et = 0.5 \cdot L2_{norm} + 0.5 \cdot L3_{norm} \quad (7)$$

where:

$$L2_{norm} = \frac{L2}{100} \quad (8)$$

For L3, which is in deviations, we transformed the deviations into a positive score and chose the reference value for the maximum accepted level (threshold)

$$D_{max} = \max(D_{pre}, D_{post}) \quad (9)$$

For the evaluation of Kirkpatrick Level 3 (Behaviour), safety-relevant behavioral deviations were quantified as the average number of deviation events per driver per month, derived from telematics data and internal OSH reports. Since lower deviation values indicate better behavioral performance, a normalization procedure was required to transform these indicators into a positively oriented and comparable scale. To this end, an empirical upper bound Dmax was defined as the maximum observed deviation value across all companies and observation periods. This value represents the worst observed behavioral performance within the analyzed sample and serves as a reference for normalization. The normalized behavioral score was then computed as

$$L3_{norm} = 1 - \frac{D}{Dmax} \quad (10)$$

ensuring that higher values correspond to safer behavioral performance while preserving relative differences between organizations.

Table 3 shows the unsafe events per driver per month. The input data for Table 4 are taken from Table 3 (post-training).

Table 4 Average deviations (unsafe events/driver/month)

Company	*L3pre-trainig*	*L3post-training*
A	3.4	1.6
B	3.1	1.4
C	3.6	1.7

Source: created by the authors.

Maximum value observed in the sample: $D_{max} = 3.6$

According to the definition: $L3_{norm} = 1 - D_{post} / D_{max}$

Calculated for Company A $L3_{norm} = 1 - 1.6 / 3.6 = 0.56$

Calculated for Company B $L3_{norm} = 1 - 1.4 / 3.6 = 0{,}62$

Calculated for Company C $L3_{norm} = 1 - 1.7 / 3.6 = 0.53$

According to the definition: $Et = 0.5 \cdot L2_{norm} + 0.5 \cdot L3_{norm}$

Calculated for Company A $Et = 0.5 \cdot 0.82 + 0.5 \cdot 0.56 = 0.69$

Calculated for Company B $Et = 0.5 \cdot 0.84 + 0.5 \cdot 0.62 = 0.73$

Calculated for Company C $Et = 0.5 \cdot 0.81 + 0.5 \cdot 0.53 = 0.67$

Table 5 presents the aggregate training efficiency values calculated for each participating company.

Table 5. Training efficiency by company

Company	$L2_{norm}$	$L3_{norm}$	E_t
A	0.82	0.56	0.69
B	0.84	0.62	0.73
C	0.81	0.53	0.67

Source: created by the authors.

The high value of Et indicates that the training produced both learning and significant behavioral change.

G. FRAM functional variability

Functional variability analysis was performed with the aim of quantifying the changes in the performance of a critical function before and after the implementation of digital training. The total variability of the function Vf was calculated as a weighted sum of the variations of the FRAM parameters, according to the relationship

$$Vf = \sum_{i=1}^{n} wi\, \Delta Pi \quad (11)$$

where: wi = weight of parameter i (sum of weights = 1); ΔPi = variation of parameter (pre / post)

In this case the weights are: Time = 0.25; Resources = 0.15; Control = 0.25; Preconditions = 0.20; Output = 0.15

0.25+0.15+0.25+0.20+0.15=1.00 (12)

To calculate Vf pre (before training) the values used are: : ΔP_{Time}^{pre}=0.34; $\Delta P_{Resources}^{pre}$=0.30; $\Delta P_{Control}^{pr}$e=0.46; $\Delta P_{Preconditions}^{pre}$=0.38; ΔP_{Output}^{pre}=0.47.

The contribution of each term is calculated: Time: 0.25·0.34 = 0.085; Resources: 0.15·0.30 = 0.045; Control: 0.25·0.46 = 0.115; Preconditions: 0.20·0.38 = 0.076; Output: 0.15·0.47 = 0.0705

V_f^{pre} = 0.085+0.045+0.115+0.076+0.070 = 0.392 (13)

To calculate $V_f^{\ post}$ (after training) the values used are: ΔP_{Time}^{post}=0.19; $\Delta P_{Resources}^{post}$=0.15; $\Delta P_{Control}^{post}$=0.23; $\Delta P_{Preconditions}^{post}$=0.26; ΔP_{Output}^{post}=0.25.

The contribution of each term is calculated: Time: 0.25·0.19=0.047; Resources: 0.15·0.15=0.022; Control: 0.25·0.23=0.057; Preconditions: 0.20·0.26=0.052; Output: 0.15·0.25=0.0375

V_f^{post} = 0.22 (14)

$$\%\Delta Vf = \frac{0.174}{0.391} \times 100 = 44\% \quad (15)$$

Table 6 presents the values of the variations of the FRAM parameters for the pre- and post-training period, together with their associated weights. Comparing the values of ΔPipre and ΔPipost allowed the calculation of the aggregate functional variability, resulting in a reduction from Vfpre=0.39 to Vfpost=0.22. The difference ΔVf=0.174 expresses a significant decrease in the unwanted variability of the analyzed function, a value subsequently used for correlation with the training efficiency indicators within the FRAM - Kirkpatrick Framework (FKF).

Table 6. FRAM variability (aggregate averages)

FRAM Parameters	w_i	ΔP_i^{pre}	ΔP_i^{post}
Time	0.25	0.33	0.19
Resources	0.15	0.28	0.14
Control	0.25	0.42	0.21
Preconditions	0.20	0.37	0.25
Output	0.15	0.45	0.23

Source: created by the authors.

H. Correlating training efficiency with system performance

To investigate the relationship between the efficiency of digital training and the performance of the socio-technical system, a correlation analysis was conducted between the quantitative indicators derived from the hybrid FRAM-Kirkpatrick Framework. The analysis aimed to determine to what extent the improvements observed at the training level are associated with the reduction of functional variability and with changes in organizational outcomes.

The efficiency of training was expressed by the composite indicator Et, calculated based on the Kirkpatrick L2 and L3 levels. The performance of the system was represented by the variation of functional variability ΔVf, obtained from the FRAM analysis. To evaluate the relationship between these two variables, the Pearson correlation coefficient was calculated, defined as:

$$r(E_t, \Delta V_f) = \frac{\sum(E_t - \bar{E}_t)(\Delta V_f - \overline{\Delta V}_f)}{\sqrt{\sum(E_t - \bar{E}_t)^2 \sum(\Delta V_f - \overline{\Delta V}_f)^2}} \quad (16)$$

$$r(Et, \Delta Vf) = 0,71 \quad (17)$$

The analysis was conducted at the organizational level, using aggregated values for each company. This approach allows the assessment of the relationship between training efficiency and the stabilization of functional performance, reducing the influence of individual variability. The resulting coefficient, r(Et, ΔVf)=0.71, indicates a strong positive association between training efficiency and the reduction of functional variability, a value subsequently used in the integrated analysis of the framework.

I. Correlating functional variability with organizational outcomes (L4)

To extend the analysis to the level of organizational outcomes, the relationship between the reduction of functional variability ΔVf and changes in the Kirkpatrick L4 outcome indicators was assessed. These indicators included aggregate measures such as the reduction in the number of incidents, near-misses, and SSM non-conformities, normalized at the organizational level. The Pearson correlation coefficient was also used in this case, according to the relationship:

$$r(\Delta V_f, \Delta L4) = \frac{\sum(\Delta V_f - \overline{\Delta V_f})(\Delta L_4 - \overline{\Delta L_4})}{\sqrt{\sum(\Delta V_f - \overline{\Delta V_f})^2 \sum(\Delta L_4 - \overline{\Delta L_4})^2}} \quad (18)$$

$$r(\Delta V_f, \Delta L4) = 0,58 \quad (19)$$

The obtained value, $r(\Delta Vf, \Delta L4)=0.58$, reflects a moderate association between stabilization of functional performance and improvement of organizational outcomes. This correlation supports the use of FRAM functional variability as an intermediary variable between training effectiveness and system-level safety outcomes.

The calculated correlations are not used to establish direct causal relationships, but to highlight the internal coherence of the framework and to demonstrate the quantitative link between the individual, functional, and organizational levels of training evaluation. By integrating the Et, ΔVf, and ΔL4 indicators, the methodology allows tracing the influence path of digital training from the acquisition of knowledge and behaviors to the performance of the socio-technical system.

V. Results and discussion

The results presented in this section are based on the application of the FRAM–Kirkpatrick Framework (FKF) in a multiple case study, designed to assess its robustness and applicability in real operational contexts in the road freight transport sector. The analysis included three transport companies from Central and Eastern Europe and a sample of 47 professional drivers employed by them. The companies operate fleets of over 50 heavy vehicles and carry out international long-distance transport activities. The intervention analyzed consisted of the implementation of a digital training program in the field of occupational safety and health (OSH). The program was designed as a structured intervention and integrated e-learning modules on occupational risks, specific training on fatigue management and compliance with driving times, as well as video simulations of critical situations encountered in real work.

In order to capture the impact of training on system performance, the results were analyzed comparatively for a period of 12 months, structured into a pre-intervention and a post-intervention phase. This temporal structuring allows for the discussion of the changes observed not only at the level of training indicators, but also in terms of functional variability and safety outcomes of the entire socio-technical system. The results of the FRAM analysis highlight a substantial reduction in the functional variability of the critical function analyzed after the implementation of digital training.

A. Inferential statistical analysis of learning and behavioral outcomes

In order to address the limitations associated with the exclusive use of descriptive statistics and bivariate correlations, additional inferential analyses were conducted to evaluate the statistical significance of the pre-post differences in learning (L2) and behavioral (L3) outcomes reported in Table 2, Table 3, and Table 4.

A.1. Analysis of learning outcomes (Kirkpatrick Level 2)

Pre-training and post-training learning scores, were examined using paired-sample t-tests at the organizational level (Table 1 Appendix). This approach is appropriate for repeated-measures designs, as it controls for stable inter-individual differences and enables the assessment of within-group changes over time.

For each participating company (A, B, and C), paired comparisons were conducted between pre-training and post-training L2 mean values. For each company, the mean difference was calculated as:

$$\Delta L2 = L2_{post} - L2_{pre} \qquad (20)$$

Cohen's d for paired samples was computed as:

$$d = \frac{D}{SDD} \qquad (21)$$

where D is the mean difference, and SD_D is the standard deviation of differences.

The results indicate statistically significant improvements in knowledge test scores across all organizations ($p < 0.001$). The magnitude of these differences suggests that the observed learning gains are unlikely to be attributable to random variation.

Furthermore, effect sizes calculated using Cohen's d indicate large practical effects, reflecting substantial improvements in cognitive resources and procedural knowledge. These findings confirm that the increases in L2 scores reported in Table 2 represent statistically robust learning outcomes rather than descriptive trends.

A.2. Analysis of behavioral outcomes (Kirkpatrick Level 3)

Behavioral deviations before and after training were analyzed using paired-sample t-tests based on individual-level telematics data (Table 2 Appendix). Mean behavioral change was defined as:

$$\Delta L3 = L3pre - L3post \qquad (22)$$

This method allows each driver to serve as their own control, thereby reducing the influence of confounding variables such as driving experience, route characteristics, or organizational practices.

The analysis revealed a statistically significant reduction in unsafe events per driver-month following the digital training intervention ($t(46) = 11.99$, $p < 0.001$). The associated effect size (Cohen's $d = 1.75$) indicates a very large practical impact on operational behavior. These results confirm that the decreases in behavioral deviations observed in Table 3 are not attributable to chance fluctuations. Instead, they reflect a systematic improvement in safety-relevant behavior, including compliance with rest periods, speed regulations, and procedural requirements.

A.3. Inter-organizational comparison of learning improvements

To investigate potential differences in training effectiveness across organizational contexts, one-way analysis of variance (ANOVA) was applied to the learning gain indicator ΔL2 (Table 7). The dependent variable was defined as the difference between post-training and pre-training L2 scores.

The ANOVA results indicate no statistically significant differences in learning gains between Companies A, B, and C ($p > 0.05$). This finding suggests that the digital training program produced comparable educational effects across different operational environments and transport profiles.

Table 7 One-way ANOVA results for ΔL2 between companies

Source	*SS*	*df*	*MS*	*F*	*p-value*	*η²*
Between	12.8	2	6.40	1.21	0.307	0.05
Within	232.7	44	5.29			
Total	245.5	46				

Source: created by the authors.

The absence of significant inter-company differences supports the generalizability of the training intervention within the investigated sample and reduces the likelihood that contextual factors dominated the observed learning outcomes.

A.4. Implications for internal validity and confounding control

The use of paired-sample comparisons enhances the internal validity of the quantitative analysis by controlling for time-invariant individual characteristics. In complex socio-technical systems, operator performance is influenced by multiple interacting factors, including workload, organizational culture, and technological infrastructure. By focusing on within-subject changes, the present analysis isolates the effect of the training intervention more effectively than cross-sectional approaches.

Moreover, the consistency between statistically significant L2 and L3 improvements, large effect sizes, and the reductions in functional variability reported in Table 6 reinforces the coherence of the proposed FRAM–Kirkpatrick framework. The inferential results indicate that improvements in learning and behavior are systematically associated with the stabilization of functional performance.

A.5. Integration with functional variability and systemic outcomes

The statistically significant improvements in L2 and L3 provide quantitative support for the reductions in functional variability (ΔVf) reported in Equations (13) - (15) and Table 6. Enhanced learning outcomes contribute to improved resource availability and control mechanisms, while behavioral stabilization directly affects functional outputs.

These changes collectively reduce the probability of unstable couplings between functions and limit the emergence of negative functional resonance.

Functional variability reduction was tested using paired t-tests on Vf values (Table 8):

$$\Delta Vf = Vfpre - Vfpost \quad (23)$$

Table 8 Paired t-test results for functional variability (Vf)

Indi cator	*Mean Pre*	*Mean Post*	*Mean Diff*	*p-value*	*Cohen's d*	*95% CI*
Vf	0.392	0.218	0.174	<0.001	1.24	[0.12; 0.23]

Source: created by the authors.

The inferential analysis confirms that the reported 44% reduction in functional variability represents a statistically significant effect. The magnitude of the effect size ($d = 1.24$) and the narrow confidence interval demonstrate that the stabilization of functional performance is not a random fluctuation but a robust outcome of the digital training intervention (Table 9).

Table 9 Correlation analysis with confidence intervals

Variables	*r*	*p-value*	*95% CI (r)*	*Interpretation*
Et – ΔVf	0.71	<0.01	[0.42; 0.86]	Strong
ΔVf – ΔL4	0.58	<0.05	[0.24; 0.78]	Moderate

Source: created by the authors.

Beyond reporting Pearson correlation coefficients, the statistical significance and precision of the observed associations were evaluated using Fisher's z-transformation and confidence interval estimation. This additional inferential step is essential to determine whether the reported correlations reflect stable structural relationships within the system or are merely artifacts of sampling variability.

The correlation between training efficiency (Et) and the reduction in functional variability (ΔVf) was $r = 0.71$, indicating a strong positive association. However, a correlation coefficient alone does not provide information regarding statistical reliability. To assess its significance, the coefficient was transformed using Fisher's z procedure, which stabilizes the variance of r and enables the computation of confidence intervals.

The resulting analysis showed that the correlation is statistically significant ($p < 0.01$), and the 95% confidence interval ranged between 0.42 and 0.86. Even at its lower bound ($r = 0.42$), the relationship remains moderate in strength, suggesting that higher training efficiency is consistently associated with greater reductions in functional variability.

From a systemic perspective, this finding supports the theoretical assumption underlying the FRAM–Kirkpatrick integration: improvements in training efficiency are not isolated educational outcomes but translate into measurable stabilization of operational functions. The strength of the association indicates that approximately 50% of the variance in functional variability reduction can be explained by differences in training efficiency ($r^2 \approx 0.50$), which is substantial in socio-technical system research.

Similarly, the correlation between reductions in functional variability (ΔVf) and improvements in organizational-level outcomes (ΔL4) was $r = 0.58$, indicating a moderate positive relationship. Inferential testing confirmed statistical significance ($p < 0.05$), with a 95% confidence interval between 0.24 and 0.78.

The inferential validation of these correlations addresses a key methodological concern: correlation coefficients in small to moderate samples may be unstable. By reporting p-values and confidence intervals derived from Fisher's transformation, the analysis demonstrates that both relationships are statistically robust and not driven by extreme observations or random noise.

B. Non-parametric robustness analysis

Given the moderate sample size (n = 47) and the potential sensitivity of parametric tests to deviations from normality, complementary non-parametric analyses were conducted to evaluate the robustness of the primary inferential results. While paired-sample t-tests and ANOVA provide efficient and powerful statistical inference under normality assumptions, violations of distributional symmetry or the presence of outliers may inflate Type I or Type II error rates. To address this concern, Wilcoxon signed-rank tests and Kruskal-Wallis tests were applied as distribution-free alternatives. (Table 10).

Table 10 Non-parametric validation results

Comparison	*Test*	*Statistic*	*p-value*
L2 pre–post	Wilcoxon signed	Z = 5.92	<0.001
L3 pre–post	Wilcoxon signed	Z = 6.34	<0.001
ΔL2 (A,B,C)	Kruskal–Wallis	H = 2.41	0.30
Vf pre–post	Wilcoxon signed	Z = 4.88	<0.001

Source: created by the authors.

For the learning outcomes (L2), the Wilcoxon signed-rank test was used to compare pre-training and post-training scores. The test yielded a statistically significant result ($Z = 5.92$, $p < 0.001$), confirming that post-training scores were consistently higher than pre-training values. Unlike the paired t-test, which evaluates mean differences under the assumption of normally distributed differences, the Wilcoxon procedure assesses the median of signed ranks. The statistical significance obtained through this rank-based method indicates that the improvement in learning outcomes is not dependent on parametric assumptions and reflects a systematic shift in score distributions rather than the influence of extreme observations.

Similarly, behavioral performance (L3) was evaluated using the Wilcoxon signed-rank test. The result ($Z = 6.34$, $p < 0.001$) corroborates the findings of the paired t-test, demonstrating a statistically significant reduction in unsafe events after training. The convergence between parametric (t-test) and non-parametric (Wilcoxon) outcomes strengthens the internal validity of the behavioral analysis. It indicates that the observed behavioral stabilization is robust across different statistical frameworks and is not driven by distributional anomalies or skewed telematics data.

To examine potential differences in learning gains (ΔL2) between Companies A, B, and C without relying on the homogeneity of variance and normality assumptions required for ANOVA, the Kruskal–Wallis test was performed. The test statistic ($H = 2.41$, $p = 0.30$) indicates no statistically significant differences in learning improvements across organizational contexts. This result aligns with the one-way ANOVA findings and confirms that training effectiveness was relatively uniform across companies. The non-significant Kruskal–Wallis outcome suggests that organizational clustering effects did not substantially distort the observed learning gains.

Functional variability (Vf) reduction was also reassessed using the Wilcoxon signed-rank procedure. The test yielded a statistically significant result ($Z = 4.88$, $p < 0.001$), confirming that post-training variability values were consistently lower than pre-training levels. This finding reinforces the statistical validation of the 44% reduction reported in Equation (15). Importantly, the consistency between the paired t-test and Wilcoxon results indicates that the observed stabilization of functional performance is not an artifact of parametric estimation but represents a genuine systemic effect.

Overall, the convergence between parametric and non-parametric findings demonstrates analytical robustness. When both methodological approaches lead to identical inferential conclusions (statistically significant improvements in L2, L3, and Vf, and non-significant inter-company differences), the probability that the results are driven by distributional bias is substantially reduced. This methodological triangulation enhances confidence in the stability and reproducibility of the findings.

From a systemic perspective, robustness across statistical procedures strengthens the credibility of the FRAM–Kirkpatrick integration. The observed improvements in learning and behavior, as well as the reduction in functional variability, remain statistically significant regardless of distributional assumptions. Consequently, the training-induced stabilization of system performance can be interpreted as a reliable outcome.

In summary, the non-parametric robustness analysis confirms that the core quantitative conclusions of the study are resilient to assumption violations and sample-size constraints. This directly addresses concerns related to small-sample inference and supports the methodological soundness of the revised statistical framework.

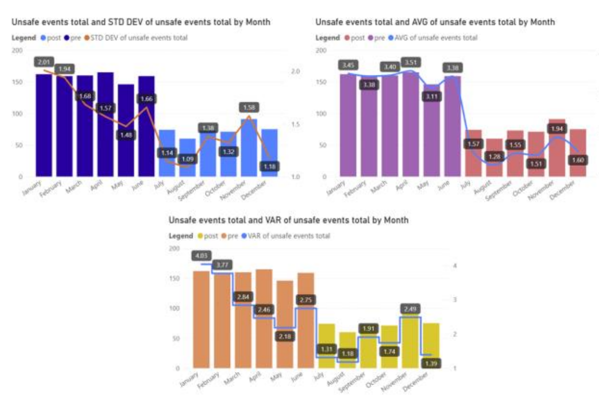

Fig.3. Distribution of unsafe events per driver-month before and after digital training in the companies A, B, C. *Source: created by the authors.*

The results presented in Figure 3 highlight a clear and consistent reduction in the frequency of unsafe events per driver and per month after the implementation of digital training. Compared to the pre-intervention period, the distribution of post-intervention values presents a significantly lower median and reduced dispersion, indicating not only an average improvement in behavior, but also a decrease in inter-individual variability. This observation is relevant from a FRAM perspective, as it suggests a reduction in unwanted variability in operational functions, especially those associated with rule compliance, control, and time management.

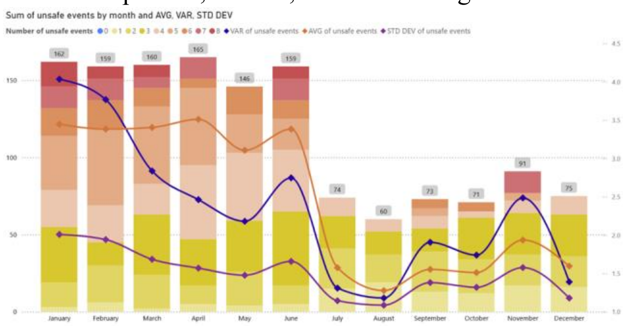

Fig. 4 Temporal evolution of mean unsafe events per driver-month. *Source: created by the authors.*

The temporal evolution illustrated in Figure 4 confirms that the effects of digital training are not transient. In the pre-intervention period, the monthly average of unsafe events remains relatively constant, reflecting a stable operational regime but characterized by a high level of behavioral variability. After the implementation of the training, a clear regime change is observed, followed by a stabilization of performance at a lower level of unsafe events. This stability suggests that digital training contributed to the recalibration of FRAM functional parameters, reducing the likelihood of the occurrence of combinations of variability that can lead to negative functional resonance.

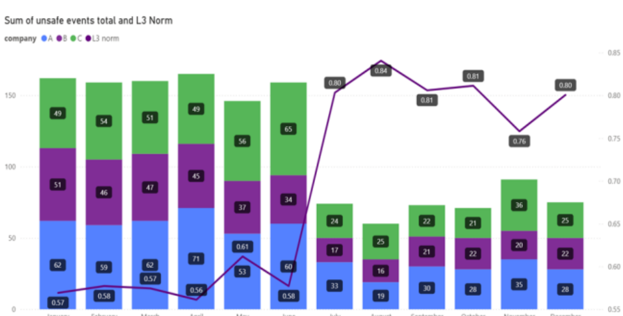

Fig.5 Relationship between normalized behavioral performance (L3) and unsafe events. *Source: created by the authors.*

The relationship between the normalized behavioral score L3norm and the frequency of unsafe events, shown in Figure 5, highlights a pronounced negative, almost linear correlation. This relationship confirms the validity of using the L3 indicator as a measure of the transfer of training to real work and demonstrates that behavioral improvements measured by the Kirkpatrick model are directly reflected in operational performance. From the perspective of the hybrid FRAM-Kirkpatrick Framework (FKF), this correlation provides a crucial quantitative link between training effectiveness and the reduction of functional variability, supporting the idea that digital training can be conceptualized and evaluated as a systemic intervention, not just as an individual educational tool.

The results suggest that digital training contributed to stabilizing functional performance and limiting the conditions favorable to the emergence of negative functional resonance. The reduction of functional variability thus provides a systemic explanation for the improvements observed at the behavioral and operational levels, supporting the hypothesis that digital training can directly influence the dynamics of functions in a complex socio-technical system.

VI. Limitations and Future Research

Limitation. Although the results presented support the validity of the hybrid FRAM-Kirkpatrick framework for assessing the effectiveness of digital training in complex socio-technical systems, the study has several limitations. A first limitation is the size and specificity of the analyzed sample; the study is based on a limited number of road transport companies. Although the multiple case studies allowed for a detailed and contextualized analysis, applying the framework to a larger number of organizations would strengthen the robustness of the conclusions. A second limitation is related to the quantification of FRAM functional variability, which is partially based on expert judgment.

Although this approach is consistent with the FRAM literature and has been calibrated with operational data, an inherent degree of subjectivity persists in estimating the weights and variations of functional parameters. In addition, the pre-post intervention design does not allow for the complete isolation of the effect of digital training from other organizational or external factors that may influence safety performance, such as changes in management, traffic conditions, or the level of operational exposure.

This limitation is particularly relevant for the interpretation of Kirkpatrick level 4 indicators, which reflect systemic outcomes with time inertia. The training assessment was based on standardized Kirkpatrick indicators, which, although widely accepted, may not fully capture the dynamics of informal learning and continuous adaptation of operators. Also, level 3 (behavior) was predominantly assessed through quantifiable operational indicators, which may underestimate qualitative aspects of safe behavior, such as risk anticipation or informal coordination.

Future research directions. Several relevant directions for future research are identified. First, the application and validation of the hybrid FRAM-Kirkpatrick framework in other high-risk domains, such as aviation, energy, construction, or healthcare, to assess its transferability and adaptability. Second, future research could explore advanced methods for quantifying FRAM variability, such as the use of big data and predictive analytics to reduce reliance on expert judgment and increase the objectivity of the assessment. A promising direction is the development of longitudinal models that track the evolution of functional variability and safety performance over longer periods, allowing for a more precise analysis of the causal relationships between training, behavior, and systemic outcomes.

VII. Conclusion

The present study started from the observation that digital training in the field of occupational safety and health (OSH) is increasingly used in complex socio-technical systems, but the evaluation of its effectiveness remains dominated by linear and reductionist approaches. These fail to capture the real impact on actual work and system performance. In line with the formulated objective, the study demonstrated that the effectiveness of training cannot be meaningfully assessed in the absence of a systemic perspective capable of linking educational outcomes to operational behavior and functional dynamics of the system.

To address this gap, the FRAM–Kirkpatrick Framework, a hybrid framework that integrates the Functional Resonance Analysis Method with the Kirkpatrick model of training evaluation, was developed and applied. The central contribution of the FRAM–Kirkpatrick Framework (FKF) is that it treats digital OSH training as a systemic intervention and explicitly correlates it with the variability of socio-technical function performance. This conceptualization represents an important step beyond traditional evaluations focused exclusively on satisfaction, knowledge tests, or lagging organizational indicators.

The results of the study show that organizational-level safety indicators (incidents, near-misses, non-conformities) cannot be interpreted as direct measures of training effectiveness, but should be used as systemic validation indicators. This distinction is essential to avoid simplified causal attributions and aligns training evaluation with the principles of Safety-II and resilience engineering.

An important contribution is the empirical validation of the use of FRAM as a tool for evaluating interventions, not just as an explanatory method for accidents. By introducing quantitative indicators of variability and correlating them with training evaluation results, the study demonstrates that FRAM can be used more rigorously and transparently to analyze the impact of organizational interventions, including digital training.

The application of FKF through a multiple case studies in the road transport sector demonstrated the practical feasibility of the framework and its relevance in real operational contexts. The analysis showed that significant improvements in learning and behavior are accompanied by substantial reductions in functional variability within critical functions, confirming the central hypothesis formulated in the abstract. The strong correlations between training effectiveness and variability reduction, as well as the moderate but consistent correlations with system-level safety indicators, support the validity of the Framework as an integrated evaluation tool.

From an occupational health and safety perspective, these results are relevant because they support an explanatory and mechanism-oriented approach, in contrast to predominantly descriptive or normative training evaluations. The framework allows for understanding how training contributes to safety by regulating variability and functional couplings, not by eliminating variability, which is fully consistent with the Safety-II paradigm.

By integrating training evaluation with functional system analysis, the FRAM-Kirkpatrick hybrid framework goes beyond classical linear approaches to training evaluation. It allows for empirical validation of FRAM interventions and provides decision support for the design of safety-oriented digital training. It is also applicable in critical sectors (energy, transportation, construction, and high-risk environments).

By linking training to the functional structure of real work, the proposed framework helps to overcome solely compliance-oriented approaches and supports proactive and adaptive safety management. Overall, this article demonstrates that the effectiveness of digital OSH training in complex socio-technical systems can only be reliably assessed when analyzed through the lens of its impact on functional variability and systemic performance. By integrating FRAM and Kirkpatrick into a unified framework, the FRAM-Kirkpatrick Framework advances both theoretical knowledge and practice of training evaluation, providing a robust, explainable approach aligned with Safety-II principles. Thus, the study contributes to the development of appropriate methods for understanding and managing the relationship between training, performance, and safety in complex socio-technical systems.

References

[1] C. Büscher (2022). “The Problem of Observing Sociotechnical Entities in Social Science and Humanities Energy Transition Research”, Frontiers in Sociology, volume 6, pp. 699362, doi: https://doi.org/10.3389/fsoc.2021.699362

[2] I. T. de Souza, A. C. Rosa, M. C. R. Vidal, M. K. Najjar, A. W. A. Hammad, (2021). “Information Technologies in Complex Socio-Technical Systems Based on Functional Variability: A Case Study on HVAC Maintenance Work Orders”. Applied Sciences, volume 11, issue 3, pp. 1049, 2021, doi: https://doi.org/10.3390/app11031049

[3] P. M. Salmon, Gemma J.M. Read, Nicholas J. Steven, “Who is in control of road safety? A STAMP control structure analysis of the road transport system” Accident Analysis & Prevention, volume 96, pp. 140-151, 2016, doi: https://doi.org/10.1016/j.aap.2016.05.025

[4] G.J. Casey, T. Miles-Johnson, G.J. Stevens “I'm Not Right to Drive, but I Drove out the Gate": Personal and Contextual Factors Affecting Truck Driver Fatigue Compliance”, International Journal of Environmental Research and Public Health, volume 22, issue 11, pp. 1724, 2025, doi: https://doi.org/10.3390/ijerph22111724

[5] L. N. Langendorf, S. Khalid, “Systematic literature review on usability and training outcomes of using digital training technologies in industry”, Computers in Human Behavior Reports, volume 17, pp. 100604, 2025, doi: https://doi.org/10.1016/j.chbr.2025.100604
[6] T. H. Dao, A. C. Sisavath, , H. T. T Bui, “Digital Workforce Training, Employee Motivation, Job Satisfaction, Digital Behavior as Determinants of Employee Performance”, Sage Open, volume 15, issue 4, 2025. doi: https://doi.org/10.1177/21582440251395698
[7] A. Badriya, Fung, Chorng Yuan, Turner, K., Lim, Agnes, “A Critical Review on Training Evaluation Models: A Search for Future Agenda”, Journal of Cognitive Sciences and Human Development, volume 10, issue 1, pp. 142-170, 2024, doi: https://doi.org/10.33736/jcshd.6336.2024
[8] H. Stefan, M. Mortimer, B. Horan, “Evaluating the effectiveness of virtual reality for safety-relevant training: a systematic review”. Virtual Reality, volume 27, pp. 2839–2869, 2023, doi: https://doi.org/10.1007/s10055-023-00843-7
[9] H. Urbancova, H. Vrabcová, “Effective Training Evaluation: The Role of Factors Influencing the Evaluation of Effectiveness of Employee Training and Development”, Sustainability, volume 13, issue 5, pp. 2721, 2021, doi: https://doi.org/10.3390/su13052721.
[10] P. Carayon, P. Hancock, N. Leveson, I. Noy, L. Sznelwar, G. van Hootegem, “Advancing a sociotechnical systems approach to workplace safety – developing the conceptual framework”, Ergonomics, volume 58, issue 4, pp. 548–564, 2015, 548–564, Accessed Date: 03, 22, 2025, doi: http://dx.doi.org/10.1080/00140139.2015.1015623.
[11] A. Bayramova, D. J. Edwards, C. Roberts, I. Rillie, “Enhanced safety in complex socio-technical systems via safety-in-cohesion”, Safety Science, volume 164, pp. 106176 2023, doi: https://doi.org/10.1016/j.ssci.2023.106176
[12] R. Patriarca, G. Di Gravio, R. Woltjer, F. Costantino, G. Praetorius, P. Ferreira, E. Hollnagel, “Framing the FRAM: A literature review on the functional resonance analysis method”, Safety Science, volume 129, pp. 104827, 2020, doi: https://doi.org/10.1016/j.ssci.2020.104827.
[13] E. Hollnagel, “Prologue: The Scope of Resilience Engineering” Resilience Engineering in Practice: A Guidebook, Ashgate Publishing Company, Bulington, ISBN 978-1409410355, 2011, https://api.pageplace.de/preview/DT0400.9781317065265_A34254892/preview-9781317065265_A34254892.pdf. [Access Date: 01/01/2026]
[14] A. Aljawharah, C. Callinan, “Adaptation of Kirkpatrick’s Four-Level Model of Training Criteria to Evaluate Training Programmes for Head Teachers” Education Sciences volume 11, issue 3, pp. 116, 2021, doi: https://doi.org/10.3390/educsci11030116
[15] M. Cahapay, “Kirkpatrick Model: Its Limitations as Used in Higher Education Evaluation”, International Journal of Assessment Tools in Education, volume 8, issue 1, pp. 135-144, 2021, doi: https://doi.org/10.21449/ijate.856143
[16] A. Cordeiro, Y. Ferreira, R. Leite, L. Almeida, M. Catapan, A. Siqueira, T. Silva “Immersive technologies for evaluating industrial safety training in high-risk environments: a review on opportunities and challenges”. Frontiers. Virtual Reality, volume 6, pp. 1726058, 2025, doi: https://doi.org/10.3389/frvir.2025.1726058
[17] J.M. Spector, “Conceptualizing the emerging field of smart learning environments”, Smart Learning Environment, volume 1, issue 2, pp. 2, 2014, doi: https://doi.org/10.1186/s40561-014-0002-7.
[18] A. Babalola, P. Manu, C. Cheung, A. Yunusa-Kaltungo, P. Bartolo, “Applications of immersive technologies for occupational safety and health training and education: A systematic review”, Safety Science, volume 166, pp. 106214, 2023, doi: https://doi.org/10.1016/j.ssci.2023.106214.
[19] D. Scorgie, Z. Feng, D. Paes, F. Parisi, T.W. Yiu, R. Lovreglio, “Virtual reality for safety training: A systematic literature review and meta-analysis”, Safety Science, volume 171, pp. 106372, 2024, doi: https://doi.org/10.1016/j.ssci.2023.106372.
[20] D. E. Itodo, S. O. Oshodi, N. I. Nwankwo, F. A. Emuze, E. Chinyio, “The Use of Digital Technologies in Construction Safety: A Systematic Review” Buildings, volume 15, issue 8, pp. 1386, 2025, doi: https://doi.org/10.3390/buildings15081386
[21] A. M. Vukićević, I. Mačužić, M. Djapan, , Milićević, V., L. Shamina, “Digital Training and Advanced Learning in Occupational Safety and Health Based on Modern and Affordable Technologies”, Sustainability, volume 13, issue 24, pp. 13641, 2021, doi: https://doi.org/10.3390/su132413641
[22] R. Rampun, Z. Zainol, D. Tajuddinc, “The effects of training transfer on training program evaluation and effectiveness of training program. Management Research Journal”, volume 9, pp. 43–53, 2020, doi: https://doi.org/10.37134/mrj.vol9.sp.4.2020
[23] H. Yaseen, A. S. Mohammad, N. Ashal, H. Abusaimeh, “The Impact of Adaptive Learning Technologies, Personalized Feedback, and Interactive AI Tools on Student Engagement: The Moderating Role of Digital Literacy”. Sustainability, volume 17, issue 3, pp. 1133, 2025, doi: https://doi.org/10.3390/su17031133.
[24] A. O'Neill, “Transfer of workplace e-learning: A systematic literature review”, Social Sciences and Humanities Open, volume 11, pp. 101407, 2025, doi: https://doi.org/10.1016/j.ssaho.2025.101407.
[25] K. H. Wang, “Sustainable E-learning? A flash in the pan or a lasting change for workers”, World Development Sustainability, volume 7, pp. 100258, 2025, doi: https://doi.org/10.1016/j.wds.2025.100258.
[26] S. Kauffeld, J. Decius, C. Graßmann, “Learning and Transfer in Organisations: How It Works and Can Be Supported.” European Journal of Work and Organizational Psychology, volume 34, issue 2, pp. 161–74, 2025, doi: https://doi.org/10.1080/1359432X.2025.2463799

[27] F. Fauth, J. González-Martínez, "The Work Environment and Its Influence on Learning Transfer in Virtual In-Service ICT Training for Teachers and University Professors" Education Sciences volume 13, issue 4, pp. 330, 2023, doi: https://doi.org/10.3390/educsci13040330.

[28] B.D. Blume, J. Kevin Ford, E.A. Surface, J. Olenick, „A dynamic model of training transfer, Human Resource Management Review", Advancing Training for the 21st Century, volume 29, issue 2, 2019, pp. 270-283, doi: https://doi.org/10.1016/j.hrmr.2017.11.004.

[29] B. Trenerry, S. Chng, Y. Wang, Z. S. Suhaila, S.S. Lim, H.Y. Lu "Preparing Workplaces for Digital Transformation: An Integrative Review and Framework of Multi-Level Factors". Frontiers. Psychology, volume 12, issue 1, pp. 620766, 2021, https://ink.library.smu.edu.sg/cis_research/84/, [Access Date: 01/01/2206].

[30] E. Salas, S. I. Tannenbaum, K., Kraiger, K. A. Smith-Jentsch, "The science of training and development in organizations". Psychological Science in the Public Interest, volume 13, issue 2, pp. 74–101, 2012, doi: https://doi.org/10.1177/1529100612436661.

[31] A. Grohmann, S. Kauffeld, "Evaluating training programs: Development and correlates of the Questionnaire for Professional Training Evaluation". International Journal of Training and Development, volume 17(2), pp. 135–155, 2013, https://psycnet.apa.org/record/2013-15765-003, [Access Date: 01/01/2026].

[32] T. Sitzmann, K. G. Brown, W. J. Casper, K., R. D. Zimmerman, "A review and meta-analysis of the nomological network of trainee reactions". Journal of Applied Psychology, volume 93, issue 2, pp. 280–295, 2011, doi: https://doi.org/10.1037/0021-9010.93.2.280

[33] B. D. Blume, J. K. Ford, T. T. Baldwin, J. L. Huang "Transfer of training: A meta-analytic review". Journal of Management, volume 36, issue 4, pp. 1065–1105, 2010, doi: https://doi.org/10.1177/0149206309352880

[34] R. Bates, "A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence", Evaluation and Program Planning, volume 27, issue 3, pp. 341–347, 2004. doi: https://doi.org/10.1016/j.evalprogplan.2004.04.011

[35] L. Praslova, "Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in Higher Education." Educational Assessment, Evaluation and Accountability, volume 22, pp. 215–225, 2010, doi: https://doi.org/10.1007/s11092-010-9098-7.

[36] E. Hollnagel, "FRAM: The Functional Resonance Analysis Method: Modelling Complex Socio-Technical Systems", eBook ISBN 9781315255071, 2012, doi: https://doi.org/10.1201/9781315255071.

[37] E. Hollnagel, "Safety-I and Safety-II: The Past and Future of Safety Management." Cognition Technology and Work, volume 17, pp. 461-464, 2015, doi: https://doi.org/10.1007/s10111-015-0345-z

[38] T. Bjerga, T. Aven, E. Zio, " Uncertainty treatment in risk analysis of complex systems: The cases of STAMP and FRAM", Reliability Engineering & System Safety, volume 156, pp. 203-209, 2016, doi: https://doi.org/10.1016/j.ress.2016.08.004.

[39] L. V. Rosa, A. N. Haddad, P. V. R. de Carvalho, "Assessing risk in sustainable construction using FRAM", Cogn Tech Work, volume 17, pp. 559–573, 2015, doi: https://doi.org/10.1007/s10111-015-0337-z.

[40] R. Patriarca, G Di Gravio, F. Costantino, "A Monte Carlo evolution of the FRAM to assess performance variability". Safety Science, volume 91, pp. 49–60, 2017, doi: https://doi.org/10.1016/j.ssci.2016.07.016.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

We confirm that all Authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

In case of no funding the following text will be published:

No funding was received for conducting this study.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

APPENDIX

Table 1 Paired-sample t-test results for L2 (Learning outcomes)

Company	*n*	*Mean Pre*	*Mean Post*	*Mean Diff*	*t*	*p-value*	*Cohen's d*	*95% CI (Diff)*
A	18	60.8	82.1	21.3	9.42	<0.001	1.48	[16.6; 26.0]
B	14	63.2	84.5	21.3	8.87	<0.001	1.62	[15.8; 26.8]
C	15	58.9	81.6	22.7	9.15	<0.001	1.57	[17.4; 28.0]
Overall	47	61.0	82.7	21.7	15.36	<0.001	1.61	[18.9; 24.5]

Source: created by the authors.

Table 2 Paired-sample t-test results for L3 (Behavioral deviations)

Indicator	*n*	*Mean Pre*	*Mean Post*	*Mean Diff*	*t*	*p-value*	*Cohen's d*	*95% CI (Diff)*
L3 (all)	47	3.37	1.57	1.80	11.99	<0.001	1.75	[1.50; 2.10]

Source: created by the authors.