

Empirical Pilot Evaluation of Corporate First-Aid Training in Virtual Reality Using PROMETHEE II

Adéla Nováková & Lilia Dvořáková

Faculty of Economics, University of West Bohemia, Pilsen, Czech Republic | novakova.adela00@seznam.cz

This Publication has to be referred as: Novakova, A[dela] & Dvořáková L[ilia] (2026). Empirical Pilot Evaluation of Corporate First-Aid Training in Virtual Reality Using PROMETHEE II, Proceedings of the 37th DAAAM International Symposium, pp.xxxx-xxxx, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-xx-x, ISSN 1726-9679, Vienna, Austria
DOI: 10.2507/37th.daaam.proceedings.xxx

Abstract

The use of virtual reality (VR) in corporate training in occupational safety, risk prevention, and first aid makes it possible to combine situated learning, motor skills practice, immediate feedback, and performance recording. The main objective of this paper is to use aggregated pilot data to evaluate how employees perceive first-aid training in VR, describe the change in their subjective self-confidence, and identify measurable performance metrics. A secondary objective is to propose a multicriteria framework for evaluating and interpreting the results. The theoretical background links technology acceptance, usability, subjective workload, and training evaluation. Aggregated questionnaire data from 50 employees are analysed descriptively; the PROMETHEE II model illustrates decision-making among training formats. Mean self-confidence was 2.99 points before training and 4.22 points after training on a five-point scale. The difference was 1.23 points; statistical significance is not confirmed without verified paired data. The scenario was completed by 49 participants, and correct use of an automated external defibrillator (AED) was recorded for 37 participants. Under the selected inputs, the model preferred VR combined with physical practice and debriefing. The contribution lies in distinguishing subjective evaluation, recorded performance, and decision preferences. The results support monitoring these dimensions jointly rather than concluding that transfer to practice has been demonstrated. Limitations include the pilot sample, absence of a control group, data aggregation, and immediate post-training measurement.

Keywords: virtual reality; first-aid training; employee self-confidence; virtual reality scenario

1 Introduction

Corporate first-aid training is intended to prepare employees to make decisions and act during an emergency. Its evaluation therefore cannot be based solely on participant satisfaction. Perceived usefulness, willingness to intervene, self-confidence, performance during practice, and independent execution of a skill represent distinct outcomes. This distinction is important when introducing virtual reality (VR), which integrates a situational scenario, user interaction, and automatic recording of actions. A review of the use of head-mounted VR systems in education and training shows that their benefits depend on the target skill, instructional task design, and the method used to verify performance [1].

A systematic review of VR-based safety training highlights the importance of intervention design and insufficient monitoring of retention [2]. The cognitive-affective model of immersive learning explains that presence and immersion can influence motivation and cognitive processing, although their effects depend on task design and user characteristics [3]. For corporate training evaluation, it is therefore necessary to distinguish participant reaction, change in perceived preparedness, performance in simulation, and subsequent behaviour in practice. This distinction corresponds to the logic of the four levels of training evaluation [4].

Studies that measure practical skills are particularly relevant to first aid. A randomized comparison of face-to-face and virtual resuscitation training found different outcomes depending on the specific performance indicator [5]. A review of augmented and virtual reality in resuscitation training documents considerable diversity in devices and instructional approaches [6]. A meta-analysis of randomized studies does not provide a simple basis for claiming the general superiority of a single training modality [7]. The findings therefore cannot be directly generalized across VR without physical interaction, VR connected to a manikin, and instructor-led blended training.

The research problem lies in the mismatch among evaluation levels: positive reactions to training, confidence in one's own abilities, and correct practical performance are not equivalent outcomes. Technology acceptance can be conceptualized through perceived usefulness and perceived ease of use [8], usability through the standardized System Usability Scale (SUS) [9], and subjective workload through the NASA Task Load Index (NASA-TLX) [10]. The analysed dataset contains aggregated scale-based indicators corresponding to these areas, but not complete item-level outputs from the standardized instruments. These constructs therefore serve as a theoretical basis for interpretation while also defining requirements for subsequent measurement.

The main objective is to evaluate the available pilot indicators of training perception, change in self-confidence, and performance in VR. The paper addresses three research questions: RQ1: How does mean self-confidence differ before training and immediately after training? RQ2: How do participants rate usefulness, engagement, and other characteristics

of the training? RQ3: What information about the course of the practical scenario is provided by system metrics? A secondary objective is to propose a transparent multilevel framework and illustrate the use of PROMETHEE II when comparing training formats. The originality of the paper lies in linking three separate layers of evidence - self-assessment, scenario performance, and a managerial decision model - while explicitly defining which conclusions can and cannot be drawn from each layer.

2 Theoretical Background and Critical Literature Review

To place the pilot evaluation in the context of current knowledge, a targeted critical review of peer-reviewed literature from 2016-2026 was conducted. Attention was focused on studies published in international journals indexed in Web of Science and Scopus, particularly systematic reviews, meta-analyses, and randomized studies in immersive education, occupational safety, and cardiopulmonary resuscitation. The aim of this section is not to produce a new systematic review, but to critically identify which conclusions are relatively consistent across studies and where methodological uncertainty remains.

The general literature on immersive education shows that the mere presence of a head-mounted display does not guarantee better learning. A review of VR applications in education emphasizes that outcomes depend on the type of skill, task design, and measurement approach [1], while a systematic review of immersive VR in higher education points to substantial heterogeneity in instructional designs and evaluation procedures [12]. In safety training, meta-analytic evidence suggests benefits of VR for knowledge acquisition and retention, but also shows that long-term follow-up is used in only a subset of studies and that theoretical grounding of interventions is often limited [2]. This tension supports the model-based assumption that immersion and a sense of presence may increase motivation and cognitive engagement, while their educational effects remain conditional on instructional design and user characteristics [3].

A similarly mixed picture emerges from research in health professions education. A meta-analysis of 31 randomized studies involving 2,407 participants found improvements in immediate knowledge and some skills outcomes in favour of VR, but the certainty of evidence was predominantly low to moderate, and the included studies did not demonstrate changes in clinical practice behaviour or patient-level outcomes [11]. From the perspective of corporate training, this distinction is fundamental: positive reactions, a subjective sense of preparedness, and performance in simulation cannot by themselves be treated as evidence of transfer to the real work environment. These evaluation levels should therefore be separated in accordance with the logic of multilevel training evaluation [4].

Research on cardiopulmonary resuscitation further shows that outcomes depend on the specific form of the VR intervention. In a randomized comparison of face-to-face and VR training, the effects differed according to the performance indicator assessed [5]. Issleib et al. found that the conventional course achieved a better result for the technical indicator of no-flow time, whereas VR participants reported greater subjective learning benefit [13]. Conversely, Moll-Khosrawi et al., in a study in which VR supplemented web-based learning, demonstrated shorter no-flow time, better overall BLS performance, and greater subjective learning benefit in the VR group [14]. Systematic reviews of resuscitation training therefore do not show universal superiority of one modality; instead, they emphasize differences among devices, degree of physical interaction, type of feedback, comparator intervention, and timing of follow-up measurement [6], [7].

For corporate first aid, the trend toward linking a virtual scenario with physical practice is particularly relevant. Newer platforms designed for workplace first responders integrate immersive environments with a manikin, haptic performance, and instructor interaction [15]. A randomized study involving non-healthcare volunteers further demonstrates the importance of monitoring skill retention over a longer time horizon rather than only immediately after training [16]. Current literature therefore supports the use of VR as part of practical training, but does not provide sufficient evidence to claim automatic superiority of VR over high-quality face-to-face instruction. A research gap remains particularly in corporate settings, where employees' subjective reactions, system-based practical performance metrics, and managerial criteria for selecting a training format are only rarely monitored together. The present pilot evaluation addresses this gap.

3 Materials and Methods

3.1 Research Design and Data Basis

The study takes the form of a one-group pilot evaluation with measurements before the intervention and immediately afterwards. The retrospective secondary analysis uses anonymized aggregated results from 50 employees who completed practical first-aid training in VR. The dataset links three types of outputs: self-confidence before and after training, immediate evaluation of the training experience, and system-recorded performance indicators during the scenario. Depending on the indicator, the reporting unit is the participant, the performed action, or the system score.

All available aggregated records from the pilot sample were included in the analysis. Metadata on recruitment method, number of individuals approached, date and location of implementation, and the structure of participating companies were not available in the analytical dataset. The sample is therefore considered non-probability based, and the findings apply to the evaluated group; generalization to a broader employee population requires a follow-up study with a predefined sampling framework.

3.2 Intervention and Operationalization of Indicators

The evaluated scenario included recognition of a life-threatening event, calling for help, cardiopulmonary resuscitation, and use of an automated external defibrillator (AED). The virtual environment was connected to a sensor-equipped manikin, enabling the recording of temporal, procedural, and motor performance parameters. In this study, performance is defined as participant behaviour during the training scenario; without a separate final test, it does not represent an independent verification of autonomous competence.

From a methodological perspective, system-supported performance must be distinguished from independent execution of the task. A complete specification of the equipment, intervention duration, number of attempts, prompts, and instructor interventions is not documented in the aggregated source data. These variables therefore do not enter the explanation of differences among participants. The findings relate to the captured course of the scenario, not to the isolated effect of the virtual environment.

Self-confidence and training evaluation were expressed on five-point scales, where a higher value indicated a more favourable rating. Self-confidence was measured before the intervention and immediately afterwards. Perceived usefulness, overall rating, engagement, realism, usability, manageable workload, and expected transfer were also evaluated. The latter indicator captures participants' expectations rather than observed transfer of skills to the work environment.

The system record included scenario completion, correct AED use, time to initiation of aid, correctness of the sequence of actions, number of errors, compression quality indicators, and an overall score. Percentage indicators have different denominators and meanings: some represent proportions of participants, while others represent aggregate correctness of actions or scores calculated by the system. They are therefore not used in the analysis as interchangeable measures of competence.

Performance metrics represent objectively recorded system indicators in the sense of instrument-based recording. Their content validity, however, depends on the technical definition of the start of time measurement, the denominators of percentage scores, and the algorithm used for the overall rating. The study therefore transparently distinguishes participant-frequency indicators from proprietary percentage scores and interprets them in relation to the course of the scenario rather than as clinical certification of competence. Table 1 provides an overview of the evaluation domains, data sources, monitored indicators, and permissible boundaries of interpretation.

Evaluation domain	Data source	Indicators	Permissible interpretation
Reaction and acceptance	Post-training questionnaire	Usefulness, overall rating, engagement, realism, usability, manageable workload	Immediate subjective reaction; not evidence of skill acquisition
Perceived preparedness	Self-confidence before and after training	Pre-test mean, post-test mean, and their difference	Change in group means; not an individual effect without verified paired data
Scenario performance	VR and manikin system record	Completion, AED, time, sequence, errors, and compression parameters	Performance under scenario conditions; not automatically independent competence or transfer
Decision preferences	Illustrative expert matrix	PROMETHEE II net flow and ranking	Preference under specified weights, scores, and thresholds; not an empirical effect of alternatives

Table 1. Operationalization of evaluation domains and interpretation boundaries for pilot indicators.

3.3 Statistical Analysis

Statistical analysis includes arithmetic means, standard deviations, absolute and relative frequencies, and the difference between post-training and pre-training means. For mean self-confidence, 95% confidence intervals were calculated using the t-distribution and $n = 50$. For the proportions of scenario completion and correct AED use, 95% Wilson confidence intervals were determined. These intervals express the uncertainty of individual aggregate estimates. They do not replace an interval for paired change because the dataset does not contain the standard deviation of individual differences or the covariance between pre-test and post-test scores. Therefore, a paired t-test, p-value, and paired effect size are not reported as verified results.

3.4 Multicriteria Decision Model

The multicriteria framework has two levels. At the evaluation level, it separates participant reaction, subjective self-confidence, and practical performance; at the decision level, it extends the assessment to operational and economic criteria. The second level is demonstrated using PROMETHEE II [17] with the alternatives A1 instructor-led face-to-face training, A2 e-learning, A3 standalone VR training, A4 VR with physical practice and debriefing, and A5 real-life exercise. The scores and weights are model inputs established to demonstrate the decision-making procedure, not measured differences among five groups of employees. The resulting ranking therefore expresses preferences under the stated assumptions and does not represent a direct estimate of the effectiveness of the individual alternatives.

All eight criteria are maximized: K1 immediate effectiveness, K2 retention, K3 training safety, K4 scalability, K5 measurability, K6 compliance and auditability, K7 acceptance, and K8 cost-effectiveness. Their weights and scores on a

1-10 scale are shown in Table 2. The criteria are not treated here as independently validated constructs. For example, measurability and auditability may overlap; weighting both may favour data-rich systems.

Alternative	K1	K2	K3	K4	K5	K6	K7	K8
Weight	0.20	0.14	0.12	0.12	0.12	0.10	0.10	0.10
A1	6.5	5	4	5	3	8	7	7
A2	4.5	4	3	9	6	6	5	9
A3	7.5	7	9	7	9	6.5	7	6
A4	9	8.5	9	7	8.5	9	8	6
A5	8	7	6	3	5	8	8	4

Table 2. Model inputs for PROMETHEE II comparison of training formats; these are not measured comparative data.

For the score difference d between two alternatives, the preference $P(d)$ equals 0 for $d \leq 0.5$, $(d - 0.5)/1.5$ for $0.5 < d < 2$, and 1 for $d \geq 2$. The aggregate preference is the weighted sum of the criterion preferences. The net flow ϕ is the difference between the mean preference of an alternative over the others and the mean preference of the others over that alternative. A higher ϕ determines a better ranking. The thresholds 0.5 and 2 are model choices, not empirically established differences in educational significance.

To assess sensitivity, three deterministic scenarios were calculated in addition to the baseline weights: the weight of cost-effectiveness K8 was increased to 0.30, the weight of compliance and auditability K6 to 0.25, and the weights of effectiveness K1 and retention K2 simultaneously to 0.30 and 0.25. In each scenario, the remaining weights were rescaled proportionally so that the sum of all weights equalled one. Alternative scores and preference thresholds remained fixed. This represents sensitivity to decision-maker priorities rather than uncertainty intervals for empirical effects. A stochastic extension of PROMETHEE [18] is an option for follow-up analysis; its results are not reported here.

4 Results

4.1 Sample Characteristics

The study included 50 employees: 26 women, 22 men, and two participants who did not report gender. Twenty-seven participants (54%) had no previous VR experience, 17 (34%) reported a single prior experience, and six (12%) reported repeated experience. The predominance of participants with no or only one prior VR experience indicates that the evaluation was not based primarily on a group of regular VR users. The characteristics describe the analysed pilot sample and do not constitute a population estimate. Descriptive self-assessment results are summarized in Table 3.

Indicator	Summary value
Self-confidence before training	2.99; SD 0.69
Self-confidence after training	4.22; SD 0.43
Difference between the reported means	+1.23 points
Perceived usefulness	4.26
Overall rating	4.19
Engagement	4.17
Realism	4.12
Usability / manageable workload / expected transfer	4.08 / 4.08 / 4.08

Table 3. Descriptive self-assessment results in the pilot sample; five-point scales.

4.2 Self-Confidence and Training Evaluation

For RQ1, mean self-confidence before training was 2.99 points (SD 0.69; 95% CI 2.79-3.19) and 4.22 points after training (SD 0.43; 95% CI 4.10-4.34). The difference between the group means was 1.23 points, corresponding to 30.8% of the full four-point range of the scale. This represents a substantial descriptive shift in perceived preparedness, although its statistical significance as a paired change cannot be verified without individual difference data. For RQ2, all monitored means exceeded four points out of five; perceived usefulness was highest (4.26), followed by overall rating (4.19) and engagement (4.17). The results indicate a favourable immediate participant reaction to the training. System-recorded performance indicators, interpreted in the following section, are summarized in Table 4.

System indicator	Pilot sample result
Scenario completion	49/50 participants (98%)
Correct AED use	37/50 participants (74%)
Time to initiation of aid	34.74 s; SD 7.70
Correctness of action sequence	89.98%; SD 6.23
Number of errors	1.56; SD 1.40
Compression depth indicator	87.22%; SD 7.51
Compression rate indicator	86.34%; SD 8.57
Chest recoil indicator	87.30%; SD 7.97
Overall system score	69.18%; SD 7.52

Table 4. System-recorded performance results during the scenario.

4.3 System Performance Indicators

For RQ3, 49 of 50 participants completed the scenario (98%; 95% Wilson CI 89.5–99.6%), and correct AED use was recorded for 37 of 50 participants (74%; 95% Wilson CI 60.4–84.1%). The 24-percentage-point difference between completion and correct execution of a critical action shows that scenario completion alone is not a sufficient indicator of mastery of the procedure. Mean time to initiation of aid was 34.74 s (SD 7.70), correctness of action sequence was 89.98% (SD 6.23), and the mean number of errors was 1.56 (SD 1.40). The overall system score was 69.18% (SD 7.52); without an externally validated threshold, it is not classified as passing or failing competence.

4.4 PROMETHEE II Model Results

In the recalculated baseline model, the ranking was $A4 > A3 > A5 > A1 > A2$. Net flows were $A4 = 0.577$; $A3 = 0.232$; $A5 = -0.072$; $A1 = -0.345$; $A2 = -0.392$. With an increased weight for K8, the ranking was $A4 > A3 > A2 > A1 > A5$; with an increased weight for K6, $A4 > A3 > A5 > A1 > A2$; and with simultaneous increases in the weights of K1 and K2, $A4 > A3 > A5 > A1 > A2$. The results apply exclusively to the stated matrix, preference function, and weights; they are not estimates of the probability of training success.

5 Discussion

5.1 Interpretation of Empirical Findings

The results provide an empirical picture of three distinct evaluation domains. After training, higher mean self-confidence was recorded, participants rated the experience favourably, and system logs revealed a difference between scenario completion and correct execution of a critical action. The practical value of the study lies precisely in their joint interpretation: a company should not evaluate training only through satisfaction or completion, but through a set of non-interchangeable indicators of reaction, perceived preparedness, and performance.

The results should be interpreted in the context of the heterogeneous findings of previous studies [5], [6], [7], [13], [14]. VR can form part of differently designed interventions, and its evaluation depends on the target skill, degree of physical interaction, feedback provided, and testing conditions. In this study, the effect of the virtual environment cannot be separated from physical practice, instructor support, and task repetition. Percentage-based system scores therefore cannot be directly compared with clinical indicators from other studies without an identical algorithm and protocol. The observed difference between the high scenario completion rate and the lower proportion of correct AED use is consistent with the broader literature, which cautions against equating user acceptance or task completion with objectively verified skill [11], [16].

For corporate training management, the findings can be translated into three subsequent decisions. First, the 98% completion rate indicates that the scenario was completable for most participants under the conditions of the pilot training. Second, the 74% rate of correct AED use identifies a specific content area for targeted debriefing and repeated practice. Third, the difference between high self-ratings and the lower proportion of correct execution of a critical action supports introducing a follow-up test without prompts. A recommended corporate output is an evaluation dashboard that, in addition to participation, reports correctness of critical steps, time, error rate, compression quality, and the result of a delayed retest.

5.2 Managerial Interpretation of the Decision Model

PROMETHEE II serves to express decision preferences transparently, not to statistically verify training effectiveness. In the baseline matrix, A4 has the highest or joint-highest score on five of the eight criteria, which explains its favourable ranking. The fact that A4 remains first in the selected scenarios does not remove the subjectivity of the inputs, possible overlap among criteria, or incomplete specification of the alternatives. Retention and cost-effectiveness are assumptions in the model; they were not measured in the pilot sample. The model therefore specifies the proposed evaluation approach but does not provide a direct recommendation to purchase a particular product.

5.3 Limitations and Further Research

The main limitations are the one-group design, pilot scale, non-probability sampling, and immediate post-intervention measurement. Aggregation prevents verification of pairing, analysis of the relationship between self-confidence and performance, and control for previous experience at the individual level. Incomplete technical documentation limits the reproducibility of system scores. The findings therefore support conclusions about the analysed sample and the feasibility of multilevel evaluation, not causal claims about the superiority of VR over other forms of instruction.

A follow-up validation of the framework should use a preregistered protocol, a primary indicator of practical performance, and a separate measure of self-confidence. A pseudonymous identifier should link the pre-test, post-test, system logs, and delayed retest. For comparison with an alternative training format, block randomization by previous experience and workplace is appropriate. The analysis should report complete pairs, mean change, 95% confidence interval, test statistic, p-value, and paired effect size. A delayed test without prompts would make it possible to assess retention and reduce confusion between performance during supported practice and independent competence. For PROMETHEE, scores and weights should be obtained through an independent expert procedure, actual costs should be added, and stochastic sensitivity analysis should be performed.

6 Conclusion

The empirical results of the pilot sample show an increase in mean subjective self-confidence from 2.99 to 4.22 points, a favourable user reaction, and a high scenario completion rate. At the same time, correct AED use reached 74%, although 98% of participants completed the scenario. This difference has direct practical significance: the success of corporate training cannot be inferred solely from participation or completion, but must also include the correctness of critical actions. The proposed multilevel framework links self-assessment, system-recorded performance, and managerial decision-making. Under the stated assumptions, PROMETHEE II favours a combination of VR, physical practice, and debriefing. The study thus provides a useful basis for managing and further validating corporate training, although without a control group and individual paired data it does not demonstrate an isolated causal effect of VR or transfer to practice.

7 Declaration

Author Adéla Nováková serves as an ambassador for VR Rescue; this professional affiliation represents a potential conflict of interest and is disclosed for transparency. The study is a secondary analysis of anonymized aggregated results; the author did not work with individual records or participants' personal data. The summary data required to interpret the results are provided in Tables 3 and 4. A generative artificial intelligence tool was used for linguistic and structural editing of the text. The author verified the calculations, expert interpretation, cited sources, and final wording of the manuscript and assumes full responsibility for them.

8 Acknowledgements

This paper was supported by project SGS-2026-008 'Global Challenges and Risk Aspects: A Financial and Accounting Perspective' at the Faculty of Economics, University of West Bohemia.

9 References

- [1] Jensen, L. & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training, *Education and Information Technologies*, Vol. 23, No. 4, pp.1515-1529, DOI: 10.1007/s10639-017-9676-0
- [2] Scorgie, D.; Feng, Z.; Paes, D.; Parisi, F.; Yiu, T. W. & Lovreglio, R. (2024). Virtual reality for safety training: A systematic literature review and meta-analysis, *Safety Science*, Vol. 171, 106372, DOI: 10.1016/j.ssci.2023.106372
- [3] Makransky, G. & Petersen, G. B. (2021). The cognitive affective model of immersive learning: A theoretical research-based model of learning in immersive virtual reality, *Educational Psychology Review*, Vol. 33, pp.937-958, DOI: 10.1007/s10648-020-09586-2
- [4] Kirkpatrick, J. D. & Kirkpatrick, W. K. (2016). *Kirkpatrick's Four Levels of Training Evaluation*, ATD Press, Alexandria, VA, ISBN 978-1-60728-008-8
- [5] Nas, J.; Thannhauser, J.; Vart, P.; van Geuns, R. J.; Muijsers, H. E. C.; Mol, J. Q.; Aarts, G. W. A.; Konijnenberg, L. S. F.; Gommans, D. H. F.; Ahoud-Schoenmakers, S. G. A. M.; Vos, J. L.; van Royen, N.; Bonnes, J. L. & Brouwer, M. A. (2020). Effect of face-to-face vs virtual reality training on cardiopulmonary resuscitation quality: A randomized clinical trial, *JAMA Cardiology*, Vol. 5, No. 3, pp.328-335, DOI: 10.1001/jamacardio.2019.4992
- [6] Cheng, A.; Fijacko, N.; Lockey, A.; Greif, R.; Abelairas-Gomez, C.; Gosak, L. & Lin, Y. (2024). Use of augmented and virtual reality in resuscitation training: A systematic review, *Resuscitation Plus*, Vol. 18, 100643, DOI: 10.1016/j.resplu.2024.100643
- [7] Sun, R.; Wang, Y.; Wu, Q.; Wang, S.; Liu, X.; Wang, P.; He, Y. & Zheng, H. (2024). Effectiveness of virtual and augmented reality for cardiopulmonary resuscitation training: A systematic review and meta-analysis, *BMC Medical Education*, Vol. 24, 730, DOI: 10.1186/s12909-024-05720-8
- [8] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, Vol. 13, No. 3, pp.319-340, DOI: 10.2307/249008
- [9] Brooke, J. (1996). SUS: A quick and dirty usability scale, In: *Usability Evaluation in Industry*, Jordan, P. W.; Thomas, B.; McClelland, I. L. & Weerdmeester, B. (Eds.), pp.189-194, Taylor & Francis, London, ISBN 978-0-7484-0460-5
- [10] Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50, No. 9, pp.904-908, DOI: 10.1177/154193120605000909
- [11] Kyaw, B. M.; Saxena, N.; Posadzki, P.; Vseteckova, J.; Nikolaou, C. K.; George, P. P.; Divakar, U.; Masiello, I.; Kononowicz, A. A.; Zary, N. & Tudor Car, L. (2019). Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration, *Journal of Medical Internet Research*, Vol. 21, No. 1, e12959, DOI: 10.2196/12959
- [12] Radianti, J.; Majchrzak, T. A.; Fromm, J. & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda, *Computers & Education*, Vol. 147, 103778, DOI: 10.1016/j.compedu.2019.103778
- [13] Issleib, M.; Kromer, A.; Pinnschmidt, H. O.; Süß-Havemann, C. & Kubitz, J. C. (2021). Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students: a randomized controlled trial, *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, Vol. 29, 27, DOI: 10.1186/s13049-021-00836-y
- [14] Moll-Khosrawi, P.; Falb, A.; Pinnschmidt, H.; Zöllner, C. & Issleib, M. (2022). Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students during COVID-19 pandemic: a randomised controlled trial, *BMC Medical Education*, Vol. 22, 483, DOI: 10.1186/s12909-022-03533-1
- [15] Carrozzino, M. A.; Caponi, M.; Pisani, S.; Papaleo, B.; Mazzei, A.; Foddìs, R.; Bergamasco, M. & Potel, M. (2026). An Immersive Virtual Reality Platform for First Aid and Emergency Training, *IEEE Computer Graphics and Applications*, Vol. 46, No. 1, pp.107-115, DOI: 10.1109/MCG.2025.3635750
- [16] Zhang, N.; Liu, Y.-J.; Yang, C.; Zeng, P.; Gong, T.; Tao, L.; Zheng, Y. & Ye, G.-Y. (2026). Long-Term Effect of Intelligent Virtual Reality First-Aid Training on Cardiopulmonary Resuscitation Skill Proficiency, *Prehospital Emergency Care*, Vol. 30, No. 1, pp.24-30, DOI: 10.1080/10903127.2025.2465718
- [17] Brans, J. P. & Vincke, P. (1985). A preference ranking organisation method: The PROMETHEE method for multiple criteria decision-making, *Management Science*, Vol. 31, No. 6, pp.647-656, DOI: 10.1287/mnsc.31.6.647
- [18] Hyde, K. M.; Maier, H. R. & Colby, C. B. (2003). Incorporating uncertainty in the PROMETHEE MCDA method, *Journal of Multi-Criteria Decision Analysis*, Vol. 12, No. 4-5, pp.245-259, DOI: 10.1002/mcda.361