BMC Medical Education

RESEARCH | Article in Press

Monitoring and evaluating disaster management training: a qualitative study of trainers' perspectives

George Voicescu · Hamdi Lamine · Luca Ragazzoni · Marta Caviglia

Received: 28 March 2026 / Accepted: 21 September 2026
© The Author(s) 2026

Abstract

Background After action reports identify disaster management (DM) training as fundamental to disaster response. However, its impact depends on the quality of the training monitoring and evaluation (M&E). Current M&E practices in DM are poorly standardized and often superficial, giving limited insight into operational readiness and a false sense of preparedness among trainees.

This study explores current M&E practices among DM training organizers, identifies perceived needs, gaps and provides a bottom-up perspective on associated challenges, and opportunities.

Methods A prospective, qualitative study employing semi-structured online interviews was used. Deliberately sampled DM training organizers were quired to understand their current M&E methodologies and the perceived needs and gaps. The audio recorded interviews were thematically analyzed, with transcripts pseudonymized and coded accordingly.

Results 53 training organizers were contacted, with 15 participating, covering diverse regions and operational contexts. The training development process, format and content were heterogeneous. Both M&E were valued, but evaluation was often prioritized, compared to monitoring, frequently using pre-post surveys and informal M&E tools. Standardized monitoring tools were lacking, while evaluation was considered superficial, not DM-specific, guided by funder requirement, rather than actual improvement. Interviewees recommend a flexible, user friendly M&E platform, dedicated to DM, with long-term tracking, follow-up, and networking capabilities.

Conclusion The heterogeneity of DM training challenges M&E, requiring adaptation to short durations, diverse topics and training methodologies. A mixed methods approach may provide a more comprehensive understanding of the learning process. The lack of monitoring might influence drastically short courses, not giving the trainers the opportunity to improve the course as it is still ongoing, leaving the participants poorly trained. As most DM trainers are operational, not educational experts, reliance on unstandardized approaches can be suboptimal, whereas dedicated M&E teams could professionalize the field. DM-adapted M&E tools could more accurately assess trainees' readiness to work, along with an M&E platform that could fill some existing gaps, while mitigating the risks associated with it.

Keywords Capacity Building standards · Educational Measurement · Disaster Medicine education · Disaster Planning · Monitoring and Evaluation · Disaster Management

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Published online: 30 September 2026

Introduction

The Sendai Framework for Disaster Risk Reduction 2015–2030 identifies training as a core component of capacity building under all 4 priority areas [1]. Moreover, the United Nations Decade of Education for Sustainable Development emphasizes the importance of monitoring and evaluation (M&E) in conducting effective assessment of the quality, relevance, and credibility of educational programs [2]. Monitoring is a continuous process that provides timely data on training progress, enabling early identification and correction of deficiencies. In contrast, evaluation is a periodic, objective assessment aimed at measuring the training's efficiency, impact, and effectiveness [3]. Disaster management (DM) is defined by the World Health Organization (WHO) as the organization, planning and application of measures preparing for, responding to and recovering from disasters [4].

Effective disaster response relies on high-quality, well-designed and efficiently delivered training, which is fundamental to building DM capacity [5–7]. Once trained, professionals are often considered operationally ready and are expected to contribute effectively throughout all phases of the disaster cycle, although most of them feel ill-prepared [8]. In this context, recent literature suggests that M&E should be regarded not merely as a training-related activity, but also as a contributor to operational effectiveness [7]. Inadequate M&E of DM trainings can create a false sense of preparedness, leaving responders vulnerable during both sudden and slow-onset disasters, impacting their willingness to work, readiness to deploy and ultimately their efficiency [9–11].

Existing literature highlights a lack of standardization in DM training programs, where evaluations are often limited to participant satisfaction and self-assessed learning improvements or feedback sessions, resulting in a heterogenous and often low-quality M&E approach that offers little insights into actual competency or operational readiness [12–16].

A recent scoping review conducted by our team revealed minimal application of existing M&E frameworks, leading towards unreliable outcomes [17]. Furthermore, the same study revealed that current M&E practices are inadequately implemented, underscoring a persistent misalignment between the stated learning objectives and the evaluation metrics employed. This inconsistency may result in inflated expectations, whereby trainees achieve high scores during evaluation despite the training program falling short of its intended outcomes. These shortcomings impact both trainees who may feel unprepared to respond to disasters, and the affected communities, who risk receiving suboptimal care [18].

There is a limited amount of data covering the field of M&E in DM, which underscores the need to better understand how these processes are currently carried out. The purpose of the present study is therefore to explore in-depth the current M&E methodologies used by the DM training organizers, map their perceived needs and gaps, and generate bottom-up understanding of the challenges and opportunities associated with M&E in DM trainings. Furthermore, the study seeks to identify the key attributes of a potential M&E platform.

Methods

Study design

A modified version of the Context, Input, Process and Product (CIPP) evaluation model served as conceptual framework, due to its decision-oriented, education-specific approach. The standard structure was adapted from its original evaluative purpose to serve as an analytical lens for examining M&E practices for the present study. (Table 1) [19] The CIPP model was used in creating the framework of the interview guide and in developing preliminary codes. An overarching hybrid deductive-inductive analysis, data driven, descriptive approach was used, involving semi structured interviews with DM training organizers. (Table 2) This design was chosen due to the low amount of existing data on the subject, impeding the usage of a more theoretical approach. The study

Table 1 The applied CIPP evaluation model

CIPP Dimensions	Collected information
Context (Goals) *What needs to be done?*	• Training objectives • M&E terminology understanding and perceived importance
Input (Plans) *How should it be done?*	• Data collection tools currently in use • Analyzed content
Process (Actions) *Is it being done?*	• Data collection practices • Training follow-up
Product (Outcomes) *Did it succeed?*	• Identified problems of existing M&E practices • Identified needs and gaps

Table 2 SPIDER research question approach for qualitative studies

Sample	Professional experienced in delivering or organizing disaster management trainings
Phenomenon of interest	Monitoring and evaluation of the disaster management trainings
Design	Overarching hybrid deductive-inductive approach
Evaluation	Personal experiences, needs and expectations regarding monitoring and evaluation of disaster management trainings
Research type	Qualitative overarching hybrid deductive-inductive approach

methods have been reported according to the Consolidated Criteria for Reporting Qualitative Research (COrEQ), the checklist being available as supplementary material (See Supplementary file 1) [20].

Participants

Participants were recruited through purposive and snowball sampling to identify trainers that had previous experience in DM educational program design, delivery, and monitoring and evaluation. The principal investigator (PI) contacted potential participants via e-mail and carefully provided information on the study objective, ethical implications, anonymity and interview methodology. Upon consent, the PI scheduled and conducted interviews with each participant, online, using the ZOOM Platform (version 6.2.7) The number of participants was determined by achieving data saturation and comprehensively answering the research questions. Patients or the public were not involved in this study.

Data collection

The interview guide was specifically developed for this study, as no suitable instrument was identified in the existing literature (see Supplementary File 2). Its design was informed by Kern et al.'s Six-step model, which served as the theoretical framework. The guide incorporated both primary and probing questions to examine training characteristics, implementation processes, goals and objectives, as well as the evaluation of existing M&E strategies [21, 22]. Secondly, the guide investigated both general and targeted needs assessment, followed by an exploration of the training organizers' expectations from an M&E platform dedicated to DM. The guide was piloted with one participant and adjusted based on feedback, then consistently applied across all interviews. Data was collected via audio-recorded online interviews conducted between March and June 2024, each lasting approximately 90 min and supplemented by detailed written notes taken during the interview.

Data analysis and reporting

Audio recordings were transcribed using the transcription feature of Microsoft® Word for Microsoft 365 MSO (Version 2504 Build 16.0.18730.20186) and subsequently verified manually. Transcripts were screened, cleaned, and pseudonymized by the PI, with all identifiable information removed, and manual notes were integrated to

enrich the dataset. Data coding and analysis followed Clarke and Braun's thematic analysis approach [23]. Two researchers independently reviewed the transcripts, applying inductive reasoning to develop preliminary codes and identify key descriptive categories emerging from participant responses. Using a collaboratively developed codebook, based on identified codes and the interview guide, transcripts were then deductively coded, and relevant quotes were extracted. Each interview required approximately four hours of analysis, carried out over five weeks using Atlas.ti software.

Ethical considerations

The study adhered to the principles of the Helsinki Declaration [24]. All participants either signed a written informed consent or provided audio-recorded consent after the study's aims and procedures were clearly explained, including their right to withdraw at any given time without negative consequence. All data were anonymized, and access to the data was restricted to the co-authors of this paper only; participant identities remained confidential and known only to the PI. This study obtained an ethical approval from the "Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca ethical review board (Ref. number "AVZ 6").

Results

53 disaster management trainers were contacted, among which 25 replied, 15 accepting to take part in the study.

All respondents were senior professionals in the field of DM training, with extensive experience ranging from 5 to 24 years. They represented both academic and operational backgrounds, working in public health institutions, international agencies, and universities, providing trainings in a wide geographical area. (Table 3) The main results of the study are summarized in Table 4.

Training characteristics

Trainees represented a broad mix of professional backgrounds, spanning both healthcare and non-healthcare fields (Table 3). Most training programs had durations between 20 min and 5 days, and 3 days being the most commonly reported length. More complex programs extended from 4 weeks to one year and were often delivered in a blended format that combined online activities with in-person sessions.

Training objectives and their development

Interviewees reported training programs covering both technical and non-technical topics. Technical content focused on prehospital triage, scene management, evacuation and prehospital care. Hospital disaster preparedness was also addressed, including topics such as the hospital safety index, disaster planning, surge capacity, rehabilitation, and recovery. Non-technical topics commonly address communication, teamwork, security, ethics, information management, cybersecurity, and interagency coordination.

Some responders reported training in educational methodologies, research in disaster settings, conflict management, decision making under uncertainty, global and public health.

Several interviewees emphasized prioritizing core principles over detailed plans. Most interviewees reported that training objectives were guided by needs assessments, typically informal and experience based. A smaller group used structured approaches such as the Six Questions method or the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model, while one responder mentioned a participatory approach, involving trainees directly in training objectives development.

Curricula were further informed by literature reviews, expert recommendations and national or international regulations and standards, (e.g., Major Incident Medical Management and Support (MIMMS), WHO Emergency

Medical Team's training standards, Bloom's taxonomy). Several responders noted that funding agencies' priorities strongly influenced the training content.

Monitoring and evaluation

Definition

a. Monitoring Responders described monitoring as an ongoing, process-oriented activity focused on the internal dynamics of training delivery. The interviewees emphasize that it encompasses assessment of content delivery, instructor performance, participant engagement, and attendance, alongside learning outcomes, such as comprehension and application of skills. Data was continuously collected through observations, assessments, participant feedback, and quantitative metrics.

b. Evaluation Training evaluation was perceived as a post-training process aimed at determining program effectiveness and justifying resource allocation. Responders emphasized assessing participants' reactions, perceived usefulness of the training, and practical application of acquired skills. Evaluation was further considered an introspective tool for trainers to identify areas for self-improvement.

Comparing training monitoring and evaluation Interviewees described monitoring and evaluation complementary, but distinct components of the training cycle. Most responders acknowledged the relevance of both processes yet reported that greater emphasis is generally placed on evaluation, as it drives accountability. One interviewee did not perceive a clear distinction between the two.

Importance

a. Monitoring Most responders described monitoring as essential for aligning training with learner needs, assessing comprehension, and identifying gaps in real time. Monitoring was deemed important to maintain adherence to schedules and curricula. Several responders emphasized the role of monitoring in financial oversight. One interviewee mentioned accessibility as a monitoring priority, explaining that it helps ensure equal access to vulnerable groups.

b. Evaluation Interviewees identified evaluation primarily to measure knowledge and skill gains, driving improvement for both trainers and trainees. One responder highlighted the formative and summative function of evaluation.

Tools and methods used

a. Monitoring Observation emerged as the core monitoring instrument. Observers were assigned to track participant actions, thought processes, and engagement during sessions. However, most teams did not use a standardized observation instrument.

Teams also used reflective techniques such as think aloud exercises and daily facilitated check-ins to promote formative feedback. One participant described these sessions as nonnegotiable moments for team reflection, sometimes held in informal settings. In longer courses, trainers relied on structured participant-mentor feedback sessions, focus groups, surveys, and confidential discussions between training directors and participants to capture cognitive and behavioral changes. Additionally, they reported engaging third-party educational experts in monitoring activities.

One interviewee described a dedicated monitoring day, when trainees meet the training leadership to discuss existing issues and propose improvements. Responders mentioned practical assessments as part of monitoring

Table 3 Profile of the interviewees

Participant ID	Years of experience in disaster management	Other Relevant Experience	Regions Worked/ Trained In	Employer	Training participants	
					Healthcare Workers	Non-Healthcare Workers
1	11	–	Europe, North America	Public University	Physicians, volunteer ambulance workers	Engineers, IT personnel
2	16	10 years conducting military training	Europe, Middle East, Asia, Africa	World Health Organization	Physicians (epidemiologist, infection prevention and control), public health professionals	Administrators, logisticians
3	10	15 years' experience in training evaluation	Europe	Public University	Different healthcare workers, without mentioning details	Civil protection staff, firefighters, public administration personnel, military, engineers, geologists
4	20	Extensive experience with United Nations and NGOs in humanitarian/development	–	Private disaster training consultancy company	Different healthcare workers, without mentioning details	Firefighters, search and rescue teams
5	17	–	Europe, North America	Public University, World Health Organization	Medical students, residents, physicians, nurses	Hospital administrators
6	15	Previous military experience	Europe, Australia, North America	Public disaster preparedness agency	Paramedics, volunteer ambulance workers	Fire department staff, mining sector workers, academic personnel
7	20	–	Europe, Africa, Asia, North America	Public University, Civil Protection agency, United Nations, UNICEF, European Union agency, ICRC, IFRC	Medical and nursing students, ambulance nurses, international Emergency Medical Teams staff	Diplomats
8	5	–	Europe, North America, Africa, Asia	International humanitarian NGO	Nurses, physicians, pharmacists, public health professionals	WASH engineers, logisticians, project managers
9	24		Asia	Public University, Public disaster preparedness agency, United Nations	—	Public disaster management responsible, disaster response volunteers
10	6	–	Asia, Africa	International humanitarian NGO, Private hospitals, Private companies	Physicians, nurses, hospital healthcare professionals	Hospital administrators, logisticians, finance staff, police, firefighters, municipal commissioners
11	7	–	Middle East, North Africa, Europe, Asia	International humanitarian NGO, Public disaster preparedness agency	Physicians, nurses, disaster response volunteers	Prison staff
12	13	Previous military experience	Europe, Asia, Africa, North America	Private disaster training consultancy company	Emergency medical care providers	Humanitarian workers (non-medical), security experts, refugee workers, trainers (training of trainers)
13	14	–	Africa, Middle East, Asia, Europe	International humanitarian NGO	Paramedics, disaster response volunteers, nurses, physicians	Fire service, police, civil protection staff, military workers, secret services, community workers

https://doi.org/10.1186/s12909-026-10465-7

https://doi.org/10.1186/s12909-026-10465-7

Table 3 (continued)

Participant ID	Years of experience in disaster management	Other Relevant Experience	Regions Worked/ Trained In	Employer	Training participants	
					Healthcare Workers	Non-Healthcare Workers
14	10	–	Middle East, North America, Europe	Public University, Public disaster preparedness agency, United Nations	Paramedics, nurses, physicians (emergency medicine)	Firefighters, police officers, hospital administrators
15	-	-	Europe, North America	World Health Organization	Clinicians, physicians (emergency medicine, epidemiology), nurses, laboratory workers	Logisticians, political scientists, emergency managers

Table 4 The main results of the study

Descriptive category		Main ideas	Quotes
Training objectives and their development		The training encompasses a wide spectrum of both technical and non-technical competencies There is limited standardization in the processes used to assess training needs Funding agencies play a substantial role in shaping the content of training programs	"*They should have a deeper insight in the challenges that major incidents or disasters come with, and to know their own plan, but also know that they need to improvise, they cannot rely on these plans"*
Definition	Monitoring	Monitoring is conceptualized as a continuous, process-oriented activity that examines internal operations and training outcomes, allowing timely modifications	*"Monitoring allows me to correct my training during the training itself."*
	Evaluation	The majority of respondents regard evaluation as a post-training activity, primarily focused on learner satisfaction and perceived usefulness A smaller subset highlights the reflective and introspective dimensions of the evaluation process	*"The evaluation is not evaluating only the participants. It's also evaluating us as trainers and to see if we did well"*
Comparing training monitoring and evaluation		Both processes are regarded as important and complementary; however, greater emphasis is placed on evaluation. One interviewee couldn't identify any difference between the two concepts	*"I think monitoring is more important than evaluation. But we live in a world that without evaluation, nothing can be done efficiently."*
Importance	Monitoring	It ensures alignment between the training and the needs assessment Early identification and remediation of gaps is promoted Promotes equitable access to training opportunities	*"Ensuring equality of access to the training programs, making sure that different skills gaps are met, disadvantaged groups are not being disadvantaged further by the training"*
	Evaluation	It assesses the skills and knowledge gain It fosters continuous improvement and adaptation of the training to the real-world contexts	*"Within my organization it would lead to change the perspective and change not only the methodology of the training, but the methodology of the capacity building itself."*
Tools in use	Monitoring	Observation is the primary method employed for data collection; however, a standardized approach is lacking Informal group discussions and individual participant–mentor sessions were also reported as common practices Monitoring is often absent in short-duration or single-session training programs	*"We do have something called one day speaks up which means I, as the director, will meet with the candidates without the trainers to fix/improve whatever problem they are facing*
	Evaluation	Written pre- and post-training surveys, along with oral discussions, emerged as the most employed evaluation methods Self-assessment is widely utilized, though it often results in conclusions of limited reliability While most respondents do not adhere to a formal evaluation framework, some referenced the use of structured approaches such as the Kirkpatrick model or the Ottawa scale Field visits are conducted as part of the evaluation process to gather data on the broader, long-term impact of the training initiatives	"*We do a reasonable job of evaluating the training in terms of, you know, did people enjoy it? Did they find it useful? Have they learned anything? I mean, almost universally people enjoy a course as long as it's not a complete disaster."*
Analyzed content	Monitoring	Operational data and resource requirements were frequently emphasized as critical considerations Mental wellbeing was deemed important The assessment of individual and group dynamics during training exercises was also noted as a relevant factor	*"If someone's stressed, they can't learn like we're trying to move away from that, completely stress people out to the point that they can't function in training"*
	Evaluation	Participant satisfaction, perceived usefulness, and knowledge acquisition were frequently cited Indicators were employed across various domains, predominantly derived from the predefined training objectives Several respondents indicated the absence of formal indicators, particularly in the context of shorter training programs	*"We have a solid foundation for the bigger programs for indicators and it's always driven by the competencies, but not for everything"*

https://doi.org/10.1186/s12909-026-10465-7

https://doi.org/10.1186/s12909-026-10465-7

Table 4 (continued)

Descriptive category	Main ideas	Quotes
Training follow-up	It is frequently absent and tends to depend on the individual trainer or the specific context When undertaken it includes efforts to assess the broader impact of the training.	*"We conducted an evaluation of the impact of the training on the mission itself and also on the final beneficiaries of the project, which is the population of the countries where the missions were deployed."*
Needs and gaps	There is a lack of standardized approaches for monitoring activities Evaluation efforts are primarily concentrated on knowledge acquisition, with limited emphasis on behavioral change or non-technical skill development M&E processes are often disconnected from operational outcomes. The inherent complexity of the DM field, including high staff turnover and unanticipated deployments, poses significant challenges to conducting long-term evaluations Concerns over potential funding loss or institutional repercussions can deter trainers from engaging in comprehensive M&E practices There is a recognized need for a user-friendly, adaptable M&E framework specifically tailored to the DM context	*"there's a lot of interesting stuff happening in the wider world of monitoring and evaluation, new techniques, new ideas, new ways of thinking which don't happen in the world of training."*
Recommendations for an M&E platform	The platform should incorporate an educational component that offers foundational knowledge on M&E It should include a technical, user-friendly feature that facilitates the development of an M&E strategy, tailored to the specific training objectives input by the user The platform should support long-term follow-up and trend analysis by enabling comparison of data across multiple training programs	*"So I would click the desired competency and you would give me the indicators to assess that"*

real-time application of skills. However, trainers reported that single-session or short-duration training often lacked monitoring.

b. Evaluation The majority of responders identified written questionnaires and verbal conversations as the main evaluation tools. Pre- and post-training evaluations focused on learning outcomes and behavioral changes assessment were frequently mentioned. Confidence and knowledge self-assessments were also widely applied.

Structured models were cited, Kirkpatrick's delayed level being mentioned.

Responders used the Ottawa scale to assess non-technical skills but consider it too lengthy for easy use. One interviewee mentioned the use of return of investment/return of expectations model and systems thinking approach.

Interviewees working within bigger organizations reported the involvement of educational specialists in item analysis and test design, maintaining databases of previous evaluation results.

Qualitative interviews and annual qualitative surveys were used to evaluate long-term training outcomes.

Hot wash-up debriefs were used for simulation evaluation, with one responder mentioning a debrief followed by a delayed evaluation to allow reflection. Some interviewees cited key performance indicators (KPIs) and exercise-specific scoring systems for simulation evaluations. Innovative evaluation methods such as color card voting and sentiment polling were mentioned.

Field based evaluation missions were used to evaluate the delayed operational impact of training programs.

Content analyzed during M&E

a. Monitoring Operational aspects such as training format, number of participants, dropout rates, test scores, time allocation were frequently mentioned. The responders mentioned tracking financial, logistical and human resources for accountability purposes.

Interviewees indicated that soft skills, students' ability to communicate, collaborate, decision-making under stress along with their mental health are frequently analyzed. The interviewees assessed the relationships between participants and instructors to understand group cohesion.

Performance during exercises was monitored from an individual, as well as group dynamics point of view.

b. Evaluation The emphasis was placed on participant satisfaction, perceived usefulness, and knowledge gain.

Practical skills, such as radio use, triage accuracy, and communication protocol adherence, emergency response center setup time, command and control indicators, surge capacity maneuvers were routinely evaluated.

In some cases, responders evaluated the whole prehospital incident management, information flow, and actors' perceptions of their roles.

Sources for these indicators derived from training objectives, needs assessment, or they were provided by funding agencies, which were noted to ask for regular updates on training progress. *"This is why it is very important to link together needs analysis and evaluation, because the needs analysis should identify the indicators that you need to use."*

Several respondents indicated that no formal indicators were in place, one interviewee stating that shorter training often lacks them.

Training follow-up

Interviewees suggested that follow-up is rarely done, being largely dependent on the trainer or the organization motivation.

A follow-up of six-months through interviews or questionnaires aimed at evaluating the usefulness and implementation of the acquired knowledge and skills was mostly cited.

One responder mentioned tracking post-deployment barriers, along with the longitudinal impact on mission and on the beneficiaries.

The main struggles of follow-up were insufficient funding, personnel turnover, or operating in unstable environments.

Needs and gaps

Several responders emphasized that M&E lacks standardization and consequently predefined benchmarks, noting that monitoring is often missing. *"It's not even standardized to have time for monitoring."*

Behavioral competencies, critical thinking, and non-technical skills were reported to be inadequately assessed. This gap, combined with evaluation methods that often stop at participant satisfaction or basic feedback, was perceived as a critical limitation. Although questionnaires were often the main M&E tools, the interviewees reported low response rates.

There was consensus that M&E results are disconnected from operational outcomes and capacity-building efforts.

Interviewees reported that the evaluation process is often outdated and fails to integrate modern M&E techniques. Complex scientific tools, while potentially valuable, were often perceived as too burdensome for field application. The interviewees emphasized the need for easy-to-use M&E tools, particularly during just-in-time training.

The responders expressed concerns about the institutional dynamics of evaluation, as they may avoid critical feedback for fear of upsetting clients. Even when evaluations are completed, they suggested that the lack of accountability prevents the implementation of recommendations.

Many respondents noted that M&E processes are deprioritized during crises when the trainers are getting deployed.

Responders' recommendations for an M&E platform

When questioned about the contents of an M&E platform, responders indicated reliable educational resources, including basic M&E information, positive and negative examples, along a glossary of specific terminology as useful to have.

Interviewees indicated that the platform should cover the entire training design cycle, from defining learning goals, designing the M&E plan and results analysis. They emphasized the need for standardized yet adaptable M&E tools repositories, templates, and sample indicators.

Several respondents further suggested enabling automated data collection and reporting, ideally integrated with existing systems.

Interviewees consistently indicated that the platform should support long-term features, like trainee tracking, trend analysis, and delayed impact evaluation. *"Something that takes away from a one-time intervention and looks at data collection over a period of time and encourages people to think in those terms."*

Usability was also considered essential, with respondents recommending role-based access, a user-friendly interface, multilingual usability, and easy adaptability to low-resource settings. Responders suggested including a space for networking and peer feedback too.

Discussion

The present study sheds light on current practice and needs in M&E in DM training by documenting how training organizers understand and apply M&E.

Interviewees profile

The interviewees possessed extensive experience in DM, offering valuable insights into the field. Their diverse geographical distribution across five continents and range of employers, from public universities and private companies to government agencies and NGOs, reflects the study's global perspective. However, although diverse experience is especially valuable in disaster management because of its international scope, contextual factors strongly shape individual perceptions. As a result, this study reflects the participants' specific experiences rather than a broader view of disaster-management training.

Training development

The heterogeneous and often informal training characteristics align with previous literature [14, 17]. Beyond healthcare professionals, many programs included non-medical staff from a wide variety of professions, reflecting the interagency nature of disaster response, which is strongly advocated by the Sendai Framework for Disaster Risk Reduction [1] This heterogeneity raises challenges for M&E, as it requires flexibility to adapt to short courses. Additionally, the M&E plan should capture role-specific outcomes, as a uniform assessment wouldn't distinguish between competencies expected of medical versus non-medical personnel. Given the high variability of these contextual factors, qualitative M&E could capture the learning progress better compared to standardized tests.

Developing training objectives in DM is challenging in the absence of widely accepted competency standards [25]. Although structured curriculum development methods were not universally applied, several respondents used established instructional design approaches that lend more validity compared with expert opinion alone [26, 27]. This gap impacts M&E, making it difficult to establish measurable outcomes and to compare different training methodologies.

The responders noted a considerable influence of the funding bodies on training objectives. In the context of M&E, trainers may prioritize meeting funders' expectations and justifying their own participation rather than assessing genuine learning improvements.

Trained topics

The wide range of training topics found in this study aligns with prior literature [15, 17]. The emphasis on soft skills underscores their increasing recognition within DM, particularly given that communication breakdowns and interagency coordination challenges are recurrent issues identified in after-action reports [28–30]. Interviewees stressed that training should include foundational disaster planning principles while remaining flexible to adapt to diverse contexts, being consistent with findings from previous studies [31, 32]. Although communication skills are more commonly assessed in DM training M&E, areas such as decision-making under stress and teamwork are rarely evaluated, as noted also by the interviewees [17]. Adapting assessment tools developed for other educational contexts could help capture changes in these soft skills more effectively [33, 34].

Monitoring and evaluation

Most interviewees described monitoring as providing timely data to inform real time adjustments, while evaluation was associated with measuring training outcomes. This distinction matches standard definitions in the M&E literature [3]. Although the definitions were not communicated ahead of the interview to the participants, the results suggest that they have a mutual understanding of the terms.

a. Monitoring

Interviewees recognized that monitoring and evaluation are complementary, yet evaluation often receives more attention, something that is also evident in the literature [17]. Monitoring, which has important internal utility for course improvement, often produces less easily quantifiable outputs and therefore attracts less external emphasis. This imbalance appears to be driven in part by funder priorities that favor measurable outcomes, which incentives trainers to prioritize evaluative reporting. The consequences of under-monitoring could be twofold. Firstly, the trainers receive meaningful feedback only at the conclusion of the training, leaving them no opportunity to refine the ongoing session. This finding is particularly relevant if discussing just-in-time training. Given their purpose, these training courses are designed to equip participants with a very precise set of skills or knowledge that they should apply immediately after the course's completion. Incorrect monitoring prevents trainers from adjusting their training in a timely manner and leaves trainees poorly prepared for field deployment.

Secondly, the lack of monitoring could promote course designs that are easily measurable, disfavoring those that might yield better learning outcomes. For example, resource allocation can be assessed through static scenarios that simply require selecting items from a list, being easily evaluated. However, dynamic simulations with shifting conditions, such as infrastructure failure, are harder to monitor but yield deeper learning about adaptive decision-making.

The responders noted using informal monitoring practices. A possible reason for the use of unstandardized approaches might be the paucity of available standardized monitoring tools, adapted to the particularities of DM [17]. Relying on informal monitoring, places the entire process on the trainer's subjective judgement, weakening the reliability of reported results. This can impact the preparedness of the trainees especially in the field of DM, where trainers are often technical, rather than educational experts, acknowledging to having limited training in M&E [35].

The reported use of education specialists was unexpected, although it represents a significant step towards the professionalization of this field. However, a lot of DM trainings have a short duration and a low budget, especially in developing, vulnerable countries where most disasters happen, the use of education specialists remaining prohibitive. Hence, flexible, easy to use standardized monitoring tools, which are suited for short training are needed to fill in this gap.

b. Evaluation

The focus of evaluations, targeting predominantly Levels 1 and 2 of the Kirkpatrick model, largely aligned with existing literature, leading to superficial evaluations [17]. There could be multiple reasons for this. Kirkpatrick's level 3 often requires long term follow-up, which is seldomly done following DM training. Assessing Level 4 is even more challenging, as disaster response outcomes are highly multifactorial and too complex to attribute to a single training intervention. Hence, strengthening and standardizing Level 3 methodologies, tailored to the specificities of DM would enhance real world behavioral changes assessment, potentially correlating better the M&E results with the actual trainee's preparedness. The limited use of indicators may reflect a scarcity of DM specific metrics, despite ongoing efforts to develop them. Existing indicators tend to prioritize technical competencies, such as triage accuracy or dispatch efficiency, while less attention is given to areas like decision-making under uncertainty or adaptability in disaster planning [36]. Using a mixed methods approach provide a deeper understanding of complex elements that are hard to evaluated otherwise [37].

The interviewees linked instructor quality to training effectiveness, but noted that evaluation was seldomly used to support instructor self-reflection and professional development [38, 39]. This observation has also been reported in more standardized programs, such as Emergency Medical Teams training initiatives, which commonly implement quality assessments through satisfaction surveys rather than validated indicators [16]. Since the instructors are not predominantly professional educators, the absence of key feedback on their training practices

constrains their professional development. Integrating M&E focused on internal improvement in the funder requirements could enhance instructional quality.

c. Follow-up

The lack of follow-up assessments, although consistent with previous studies, was contrasted by one example where follow-up evaluation missions helped assess behavioral changes and community impacts of DM trainings.

Due to their operational duties, follow-up activities are often delayed when trainers are deployed. This interruption in the M&E cycle leads to weak organizational learning and loss of critical information regarding behavioral changes and actual real-world impact of the training [16]. Additionally, the accountability in front of the funders is also impaired, because no real changes can be reported without follow up. Therefore, creating dedicated M&E teams to work during training or in follow-up missions could be a practical solution. For smaller organizations that are seldomly organizing training, outsourcing M&E activities could be a cost-effective option, at the cost of building internal capacity.

d. Existing needs and gaps

The M&E efforts are furthermore hampered by the low response rates to evaluation surveys, a challenge also noted in other educational fields. Literature suggests different strategies to minimize non response bias, based on creative and persistent follow-up strategies, providing incentives or shortening the survey can help improve participation rates [40, 41].

Advanced M&E approaches such as theory of change, design-focused evaluation, big data analytics, and machine learning are increasingly employed in education, but rarely applied in DM training [42–44]. Although complex, these conceptual frameworks might be better suited to the DM context than linear models like Kirkpatrick's model, as they are designed to handle complex, non-linear environments where outcomes depend on multiple interacting actors, assumptions, and contextual factors.

Their low usage rates may be attributed to the inherent complexity of such methods and a general lack of awareness among DM trainers regarding advanced M&E strategies. One potential solution is to provide trainers with access to standardized M&E tools, tailored to DM, that are both practical and easy to implement within their training programs.

Monitoring and evaluation platform

A platform that combine theoretical guidance and practical tools may help professionalize M&E in DM [45, 46]. Incorporating a repository of knowledge, standardized tools, validated indicators, and adaptable templates would enable trainers to identify and apply suitable approaches for their training contexts. Furthermore, integrating existing data collection and analysis software with long-term trend analysis and database creation could enhance training follow-up and continuous improvement, allowing trainers to compare different learning strategies and identify the best ones. Considering the limited awareness of M&E practices within the DM field, establishing a digital space for networking and knowledge exchange would foster peer learning, strengthening the broader training community. Beyond participants perspectives, the authors identified some broader ethical and implementation considerations for developing such platform. Risks include data misuse or misinterpretation, if standardized indicators are removed from their contextual grounding and used as proxies for competence or performance. In such cases, M&E data may be repurposed for accountability or funding decisions rather than learning, undermining formative evaluation. Finally, while standardization may enhance comparability, excessive standardization risks oversimplifying a field characterized by contextual variability and adaptive decision-making. Data protection and security are additional risks that need to be mitigated from the beginning.

Future research recommendations

Further efforts are needed to create M&E tools dedicated to DM, which could support continuous learning, enhance training quality, and contribute to the professionalization of the field. In addition, future research should prioritize the development of monitoring tools, as well as results-oriented M&E strategies capable of capturing the broader impact of training programs.

Limitations

There are several limitations associated with this study. Firstly, there is a selection bias, since only the trainers that replied to the initial email were included. Therefore, we assume them to be more interested in advancing education in DM and their perspectives may not necessarily reflect that of the broader DM trainer's population.

Secondly, given the qualitative nature of the study, the results are susceptible to recalling and reporting bias, which could over or underestimate their M&E practices. The limited number of participants may limit the contextual representation and transferability of the study's findings. Nonetheless, the interviewees reported training experiences across a wide geographical scope, teaching in a wide variety of settings, which may help to mitigate this constraint.

Additionally, we acknowledge that DM represents a broad field, encompassing a wide range of sectors, organizations, training formats, and operational contexts. Therefore, the findings may not fully capture the diversity of organizational contexts and should not be considered representative of all DM training settings.

Conclusion

DM training organizers report that M&E remain heterogeneous, evaluation being consistently prioritized over monitoring and largely limited to pre–post surveys and informal M&E tools. Standardized monitoring practices and indicators are frequently absent, particularly in short-duration trainings, while evaluation approaches tend to be superficial and insufficiently tailored to disaster-specific competencies. These findings underscore the need to reorient M&E toward meaningful assessment of real-world preparedness, as the ultimate purpose of M&E is to asses and enhance effectiveness in the field rather than to assess training outputs in isolation.

Abbreviations

M&E	Monitoring and evaluation
DM	Disaster management
CIPP	Context, input, process, product
PI	Principal investigator
COREQ	Consolidated Criteria for Reporting Qualitative Research
ADDIE	Analysis, Design, Development, Implementation, and Evaluation
MIMMS	Major Incident Medical Management and Support
KPI	Key performance Indicators

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1186/s12909-026-10465-7.

Acknowledgements This manuscript is the result of a study conducted in the framework of the International PhD in Global Health, Humanitarian Aid and Disaster Medicine organized by Università del Piemonte Orientale (UPO).

Authors' contributions GTV was the principal investigator and oversaw conceptualization, methodology development, data curation and analysis, original draft preparation, review and editing. HL oversaw conceptualization, methodology de-

velopment, data analysis, review and editing. LR oversaw conceptualization, methodology development, review and editing. MC oversaw conceptualization, methodology development, original draft preparation, review, editing and supervision.

Funding The authors did not receive support from any organization for the submitted work.

Data availability The anonymized datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate The study adhered to the principles of the Helsinki Declaration. This study obtained an ethical approval from the "Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca ethical review board (Ref. number "AVZ 6"). All participants either signed a written informed consent or provided audio-recorded consent after the study's aims and procedures were clearly explained, including their right to withdraw at any given time without negative consequence.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. UN International Strategy for DRR. Sendai Framework for Disaster Risk Reduction 2015–2030. Volume 37. UN World Conf Disaster Risk Reduct; 2015.
2. Buckler C, Creech H. Shaping the Future We Want: UN Decade of Education for Sustainable Development (2005–2014) : Final Report. Paris: UNESCO; 2014.
3. UNICEF Evaluation Office. A UNICEF guide for monitoring and evaluation : 1991 Jan 1 [cited 2022 July 3]; Available from: https://digitallibrary.un.org/record/166694
4. World Health Organization. Glossary of health emergency and disaster risk management terminology [Internet]. World Health Organization; 2020 [cited 2025 Dec 1]. Available from: https://apps.who.int/iris/handle/10665/331716
5. Kahn LH, Barondess JA. Preparing for Disaster: Response Matrices in the USA and UK. J Urban Health. 2008;85(6):910–22.
6. Yin H, He H, Arbon P, Zhu J, et al. Optimal qualifications, staffing and scope of practice for first responder nurses in disaster. J Clin Nurs. 2012;21(1–2):264–71.
7. Kaim A, Adini B, Bodas M. Evaluating the operational effectiveness of Emergency Medical Team field hospitals: A novel tool for measuring surge capacity performance. Int J Disaster Risk Reduct. 2026;132:105987.
8. Almukhlifi Y, Crowfoot G, Hutton A. Emergency Health Care Workers' Preparedness for Disaster Management: An Integrative Review. Prehospital Disaster Med. 2023;38(S1):s149–149.
9. Zagelbaum NK, Heslin KC, Stein JA, Ruzek J, et al. Factors influencing readiness to deploy in disaster response: findings from a cross-sectional survey of the Department of Veterans Affairs Disaster Emergency Medical Personnel System. BMC Emerg Med. 2014 July;19(1):16.
10. Damery S, Wilson S, Draper H, Gratus C, et al. Will the NHS continue to function in an influenza pandemic? a survey of healthcare workers in the West Midlands, UK. BMC Public Health. 2009;9(1):142.
11. Sultan MAS, Løwe Sørensen J, Carlström E, Mortelmans L, et al. Emergency Healthcare Providers' Perceptions of Preparedness and Willingness to Work during Disasters and Public Health Emergencies. Healthcare. 2020;8(4):442.

12. Ashcroft J, Byrne MHV, Brennan PA, Davies RJ. Preparing medical students for a pandemic: a systematic review of student disaster training programmes. Postgrad Med J. 2021 June;97(1148):368–79.
13. O'Malley G, Perdue T, Petracca F. A framework for outcome-level evaluation of in-service training of health care workers. Hum Resour Health. 2013;11(1):50.
14. Voicescu GT, Valente M, Della Corte F, Becerril M et al. Medical students' education in disaster medicine: A systematic literature review of existing curricula. Int J Disaster Risk Reduct. 2022 July 1;77:103090.
15. Baetzner AS, Wespi R, Hill Y, Gyllencreutz L, et al. Preparing medical first responders for crises: a systematic literature review of disaster training programs and their effectiveness. Scand J Trauma Resusc Emerg Med. 2022;30(1):76.
16. Kaim A, Amat Camacho N, von Reding H, Adini B, et al. Bridging gaps in Emergency Medical Team (EMT) capacity strengthening activities in the WHO European region. Int J Disaster Risk Reduct. 2025;131:105916.
17. Voicescu GT, Lamine H, Loșonți AE, Lupan-Mureșan EM, et al. Monitoring and evaluation in disaster management courses: a scoping review. BMC Med Educ. 2025;25(1):188.
18. Waring S, Skryabina E, Goodwin D, Lino D, et al. What components of emergency preparedness exercises improve healthcare practitioners' emergency response learning? Int J Disaster Risk Reduct. 2021;62:102357.
19. Stufflebeam D, Zhang G. The CIPP Evaluation Model: How to evaluate for improvement and accountability. 2017.
20. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care J Int Soc Qual Health Care. 2007;19(6):349–57.
21. Bordage G. Conceptual frameworks to illuminate and magnify. Med Educ. 2009;43(4):312–9.
22. Thomas PA, Kern DE, Hughes MT, Tackett SA, et al. Curriculum Development for Medical Education: A Six-Step Approach. JHU; 2022. p. 389.
23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101.
24. World Medical Association. Declaration of Helsinki ethical principles for medical research involving human subjects. JAMA J Am Med Assoc. 2013;310(20):2191–4.
25. White P, Di Benedetto F, Hansen K, Medcalf S, et al. Not Ready: The Need for a Training Standard in Healthcare Emergency Preparedness. Health Secur. 2025;23(3):147–54.
26. Pilcher J, Bedford LA. Hierarchies of evidence in education. J Contin Educ Nurs. 2011;42(8):371–7.
27. Trust T, Pektas E. Using the ADDIE Model and Universal Design for Learning Principles to Develop an Open Online Course for Teacher Professional Development. J Digit Learn Teach Educ. 2018;34:219–33.
28. National commission on terrorist attacks. The 9/11 Commission Report [Internet]. United States of America; 2004 July [cited 2025 Dec 1] p. 585. Available from: https://www.9-11commission.gov/report/911Report.pdf
29. Huang J-S, Lien Y-N. Challenges of emergency communication network for disaster response. In: 2012 IEEE International Conference on Communication Systems (ICCS) [Internet]. 2012 [cited 2025 Dec 2]. pp. 528–32. Available from: https://ieeexplore.ieee.org/document/6406204
30. Oden RVN, Militello LG, Ross KG, Lopez CE. Four Key Challenges in Disaster Response. Proc Hum Factors Ergon Soc Annu Meet. 2012 Sept;56(1):488–92.
31. Goniewicz K, Goniewicz M, Włoszczak-Szubzda A, Burkle FM, et al. The importance of pre-training gap analyses and the identification of competencies and skill requirements of medical personnel for mass casualty incidents and disaster training. BMC Public Health. 2021;21(1):114.
32. Webb GR, Chevreau F-R. Planning to improvise: the importance of creativity and flexibility in crisis response. Int J Emerg Manag. 2006;3(1):66–72.
33. Wooding EL, Gale TC, Maynard V. Evaluation of teamwork assessment tools for interprofessional simulation: a systematic literature review. J Interprof Care. 2020;34(2):162–72.
34. Salimi M, Hidayah R, Suhartono, Susiani TS, et al. Evaluating the Impact of the DECIDE Model on Decision-Making Skills in Elementary Teacher Education Students in Indonesia. J Lesson Learn Stud. 2024;7(3):449–60.
35. Hagelgans DD. Education vs. Experience: Solving the Puzzle [Internet]. 2024 [cited 2025 Dec 2]. Available from: https://www.millersville.edu/cdre/files/educationvexperience.pdf
36. Markou-Pappas N, Lamine H, Ragazzoni L, Caviglia M. Key performance indicators in pre-hospital response to disasters and mass casualty incidents: a scoping review. Eur J Trauma Emerg Surg Off Publ Eur Trauma Soc. 2024;50(5):2029–37.
37. Bamberger M, Vaessen J, Raimondo E. Dealing With Complexity in Development Evaluation: A Practical Approach. SAGE; 2015. p. 481.
38. Darling-Hammond L. Teacher Quality and Student Achievement. Educ Policy Anal Arch. 2000;8:1–1.
39. Goe L. The Link between Teacher Quality and Student Outcomes: A Research Synthesis [Internet]. National Comprehensive Center for Teacher Quality; 2007 Oct [cited 2025 Nov 4]. Available from: https://eric.ed.gov/?id=ED521219
40. Ericson A, Bonuck K, Green LA, Conry C, et al. Optimizing Survey Response Rates in Graduate Medical Education Research Studies. Fam Med. 2023;55(5):304–10.
41. Phillips AW, Reddy S, Durning SJ. Improving response rates and evaluating nonresponse bias in surveys: AMEE Guide 102. Med Teach. 2016;38(3):217–28.
42. Ersozlu Z, Taheri S, Koch I. A review of machine learning methods used for educational data. Educ Inf Technol. 2024;29(16):22125–45.

43. Smith C. Design-focused evaluation. Assess Eval High Educ. 2008;33(6):631–45.
44. Innovation US. MEL 360: Monitoring, Evaluation and Learning for Systems Change [Internet]. Medium. 2025 [cited 2025 Nov 4]. Available from: https://medium.com/@undp.innovation/mel-360-monitoring-evaluation-and-learning-for-systems-change-babc5ef9a5dd
45. Social innovation in. health: monitoring and evaluation framework [Internet]. 2021 [cited 2025 Nov 4]. Available from: https://wkc.who.int/resources/publications/i/item/social-innovation-in-health-monitoring-and-evaluation-framework
46. United Nations Development Programme. Systems Monitoring Learning & Evaluation (SMLE) [Internet]. 2024 [cited 2025 Nov 11]. Available from: https://www.undp.org/sites/g/files/zskgke326/files/2025-01/smle_resources_2024_.pdf#:~:text=Our%20initiative%20includes%20the%20M%26E%20Sandbox%2C%20an%20effort,blogs%20of%20M%26E%20Sandbox%20events%20held%20since%202023

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

George Voicescu[1,2,3] · Hamdi Lamine[4] · Luca Ragazzoni[1,3] · Marta Caviglia[1,5]

✉ George Voicescu
george.voicescu@uniupo.it

1 CRIMEDIM—Center for Research and Training in Disaster Medicine, Humanitarian Aid and Global Health, Università del Piemonte Orientale, Novara 28100, Italy

2 Emergency Medicine Discipline - Department 6 Surgery, "Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca, Romania

3 Department for Sustainable Development and Ecological Transition, Università del Piemonte Orientale, Vercelli, Italy

4 Department of Community Health Nursing, College of Nursing, University of Hail, Hail, Saudi Arabia

5 Department of Translational Medicine, Università del Piemonte Orientale, Novara, Italy