The Future of Evacuation Drills: Assessing and Enhancing Evacuee Performance

Steve Gwynne[a], Martyn Amos[b,*], Max Kinateder[c], Noureddine Bénichou[c], Karen Boyce[d], C. Natalie van der Wal[e], Enrico Ronchi[f]

[a]*Movement Strategies, UK.*
[b]*Northumbria University, UK.*
[c]*National Research Council, Canada.*
[d]*University of Ulster, UK.*
[e]*University of Leeds, UK.*
[f]*Lund University, Sweden.*

Abstract

Evacuation drills are generally the main mechanism for improving or measuring occupant performance in emergency situations, but their effectiveness is often hard to measure, and there is limited evidence for sustained training benefits. However, innovations in technology (e.g., augmented/virtual reality, novel sensors and wearable tech) offer (when combined with new approaches to designing and delivering drills) significant opportunities for a "next generation" of *evidence-based* evacuation drills. In this paper, we present the findings of a recent trans-national research project; we establish the main limitations of existing drills, propose a framework for the assessment of both training and evaluation aspects of drills, make a number of recommendations, and suggest a programme of work for their implementation. The paper, therefore, provides a conceptual foundation for future work which will focus on (1) establishing an evidence-based methodology for assessing evacuation drills (and alternatives), (2) harnessing novel objective and automatable approaches to data capture/analytics in order to better characterize performance, (3), developing alternatives to the current drill model, based on emerging technologies, and (4) developing guidance for regulatory bodies on the costs and benefits of each approach for different sce-

*Corresponding author.
Email address: martyn.amos@northumbria.ac.uk (Martyn Amos)

narios.

Keywords: Evacuation drill; protocol; training; assessment; AR/VR; simulation.

1. Introduction

An evacuation drill (ED) is a pre-planned simulation of an emergency evacuation given a specific scenario. This is conducted to assess the evacuation procedure, and to directly or indirectly improve the performance of occupants and staff involved; training benefits derived from participation or observation may also lead to procedural enhancements. Although EDs are informed by a range of safety legislation and building codes, their merits are still not well-understood (given limitations in how they are conducted and how they are observed), and their impact on evacuation performance is not well-characterized [1, 2, 3, 4, 5]. However, they are still seen by many parties as a key component of safety planning/building certification.

The availability of new approaches and technologies such as augmented/ virtual reality [6, 7, 8], computational simulation [9, 10], smart sensors/building intelligence [11] and video analytics [12] means that we now have an unprecedented opportunity to enhance the way that we plan, deliver, observe and analyze the results of evacuation drills or complementary activities, and to improve evacuation performance (and the assessment of such). This allows us to disentangle the training benefits and assessment of drill effectiveness, by following a case-control approach comparable to research practices in evidence-based medicine. The ultimate goal is to move towards a position where evidence-based evacuation drills are the norm [13, 14].

In this paper, we explore new technologies, methodologies and perspectives to (1) enhance the training component of evacuation drills, (2) improve the analysis and interpretation of their results, and (3) reduce both short-term risk to participants and operational disruption. Working towards these objectives may bring a range of beneficial outcomes, such as potentially reducing costs, improving training effectiveness and allowing more fine-grained assessment of occupant behavior.

The core novel contributions of the paper are (1) a comprehensive review of the current state-of-the-art in evacuation drills (in terms of both theoretical research and practical implementation), (2) a robust set of criteria for assessing approaches to evacuation drills, based on a community consultation exercise, and (3) a new framework for comparing the costs/benefits of different approaches against these criteria to aid the selection and practical implementation of such approaches. We base the paper on a roadmap document produced by our NEED (Networking Activities for Enhanced Evacuation Drills) project [15].

The paper is organized as follows: in Section 2 we provide the context to evacuation drills, and highlight some over-arching issues with them. In Section 3 we identify a number of specific challenges presented by EDs, in Section 4 we present our methodology for obtaining a community-sourced list of assessment criteria, in Section 5 we give the results of our consultation, and in Section 6 we show how these may be integrated into a new approach for comparing different approaches to EDs. We conclude with a discussion of the implications of our findings, and make some recommendations about possible future lines of research.

2. Background

EDs are a model of an emergency evacuation from a particular building. In turn, a model is a simplified abstraction of a real-world phenomenon. Evacuation drills are one attempt to model emergency evacuation or to reproduce conditions which could be used to train populations exposed to an emergency scenario; others include computational simulations, behavioural experiments, or questionnaires. In reality, a drill may be a model of a range of different emergency procedures, depending on the structure or scenario involved; e.g. not all emergency drills necessarily reflect evacuation to an external location. As with all models, EDs are based on a set of simplifications: much of the models credibility relies on the nature and extent of these simplifications and our understanding of their impact on the effectiveness of the model.

In the best case, an ED would be based on (1) a representative occupant population, (2) the procedural resources expected to be available during an emergency (safety staff, technology and plans), and (3) a building design that reflects route availability during the incident(s) being examined. The potential for " real-world" similarity is one of the strengths of the ED model; that is, it typically uses good proxies for the intended application domain. Where these elements are in place, the drill model might approximate real-world conditions for the scenario in question, at least while incident development and evacuee exposure to an external hazard is at a minimum. This prompts two key questions: (1) How frequently do we obtain such a close match between the real-world and the model, and (2) What are the implications of any shortfall?

Currently, *EDs are not conducted consistently*, largely due to variability in regulations, and the application of such regulations within a jurisdiction (e.g., [16]). For example, different regulatory bodies, (e.g. National Fire Protection Association, International Code Council, etc.) have different requirements for ED implementations regarding their frequency, participation, the scenario examined, oversight, and the data that needs to be recorded across different occupancies [14]. Therefore, the specific regulatory code adopted might influence both the ED requirements and their subsequent performance.

More confounding, still, is that even within a single jurisdiction, different organisations interpret the regulatory requirements differently, and apply them with varying degrees of rigour [14]. This leads to a wide variety of practices in the populations involved, the procedures employed, and building types involved, potentially undermining the effectiveness of the real-world proxies modelled [16]. Differences in jurisdictional requirements are exaggerated by variation in practices in oversight and enforcement; i.e. whether the drills are monitored and enforced to ensure performance levels.

Perhaps most fundamentally challenging is that *EDs are conducted according* *to different (and potentially competing) objectives*, often attempting to both *assess performance* and *train populations*. We contend that an ED can either be employed to assess *or* it can be employed to train. However, it cannot do *both* simultaneously without significant effort, care and resources (which are often unavailable or not applied). For example, the resources required to assess performance might include the deployment of equipment for monitoring human behaviour (or the availability of observers on site). For a combined training purpose, resources would also be required to arrange post-ED debriefing sessions or specific arrangements to enhance the educational goal of an ED. This will result in a larger number of resources/effort required (possibly both in terms of people as well as equipment). In addition, little attention is paid to the effectiveness of the drills at meeting these objectives. For instance, holding drills more frequently may have a diminishing impact of the response of the target population and at some point actually have a negative impact on performance, with people potentially associating the notification system with a drill rather than a real incident should measures not be taken [17].

Furthermore, although EDs should be observed and documented rigorously, there is often little attention paid to gathering results on key indicators, and on interrogating the impact of the EDs on performance. If EDs are documented, the quality of the data may strongly depend on the efforts performed by the data collectors in obtaining objective and reliable information. Unfortunately, EDs are frequently seen as an inconvenience (little more than a troublesome "tick-box exercise"), rather than an opportunity to gather and analyse safety-critical information [18].

These issues undermine the value of what might otherwise be a sophisticated and important model of evacuation performance, and often leads to them not being used to their full potential. Several questions, therefore, naturally arise [14]:

- How may we exploit EDs more effectively to both train and assess performance?

- What alternative models are available to both train and assess evacuee performance?
- How might we use EDs and alternative models together?

In 2018, a joint UK-Canada project was set up to investigate these questions. The NEED (Networking Activities for Enhanced Evacuation Drills) project was managed by several of the current authors, and had the over-arching goal of producing a roadmap document for the next generation of evacuation drills [15]. The project was framed by the questions above, and it involved 30 experts drawn from 5 different countries. In what follows, we summarise our methodology and main findings. We begin by considering, in detail, the main challenges presented by evacuation drills.

3. Main challenges in evacuation drills

The value of an ED is enormously dependent on the specifics of the drill execution; it is not derived simply from the fact that a drill has been performed. This becomes even more evident when we consider what we assessed are the two primary objectives of EDs: to train the evacuating population in the emergency procedure and the safety staff in their roles in the procedure during a representative scenario, and/or to assess the performance of the population, procedure and staff under the same scenario.

These are both extremely important objectives; however, we should not take it for granted that they may be achieved, nor is their achievement independent of each other. It is important to identify issues with EDs, especially those issues that may affect our ability to meet these objectives. There are significant issues with EDs, as currently executed [13, 14]: (1) the effectiveness of the ED model is not well understood; (2) drills carry both an inherent risk to participants and a significant cost (in terms of both monitoring the drill using technology and/or human and resources, and temporary loss of building functionality) [14]; (3) sub-populations are often excluded from drills (e.g., those with medical issues, or mission-critical staff), which affect the potential for training and assessment, and self-evacuation procedures are often very different to those for assisted evacuation [19]; (4) drills are conducted inconsistently across organizations, building types and between jurisdictions; (5) drills are employed to train, to assess or to do both; and (6) repeated drills may actually undermine compliance (the so-called "Cry wolf effect") [20, 17].

For an ED to be of practical use, several design questions need to be addressed. The drill designer needs to ensure that the drill conditions are reflective of real-world scenarios. This is necessary to ensure that the drill is a good enough model for the task at hand. Inherent to this is the ability for the ED designer to configure the drill, and to ensure that it captures key factors of interest. Independently of the scenario being examined, the ED should produce data in form and content that allows its performance to be assessed (i.e., the data needs to be objective and reliable). The ED should also operate at the levels of performance that would be expected during a real incident, enabling data to be both comprehensive and rich enough to provide insights. We will examine these last two points in detail, as they are most relevant to the work outlined in this article.

EDs may address performance at a number of different organizational and procedural levels. This is necessary, given that any emergency procedure is an amalgam of resources, roles, tasks, actions and objectives [21, 22]. These include recording the effectiveness of individual actions (e.g., can a specific floor warden operate a fire extinguisher?), individual roles (e.g., can a floor warden, in general, fulfil their obligations?), groups (of roles) defined within the procedure (e.g., do floor sweepers search a particular location as expected?), interactions between different groups (e.g., do security staff, wardens and buddies perform in conjunction with each other?), planned procedures (e.g., does the phased evacuation operate as expected?), interaction between different procedures (e.g., do security measures impede the evacuation plan?), interactions between external agencies (e.g., do the attending fire and paramedic services coordinate effectively?), and involvement of multiple buildings and locations (e.g., how do we manage the arrival of multiple populations at shared assembly points?)

Although the focus of this discussion is on drills conducted in buildings, we acknowledge that similar questions might be posed of other structures and other scenarios (e.g. community evacuation). Given these twin objectives of EDs, it is imperative that that both training and assessment address (in some way) these different organizational and procedural levels. In order to do so, evacuee performance needs to be understood and measured at these various levels. Similarly, different aspects of the evacuee response may be measured/influenced in a number of ways, irrespective of the scenario being examined. These performance aspects again prompt several questions: how long does it take for people to establish the reality of a situation, recall the procedural response, determine an appropriate action, identify an appropriate objective, and identify the (technological, procedural and human) resources available and the appropriate use of such resources?

Since an ED typically involves a population engaging in a procedure, it is a significant challenging to organise a drill in response to a credible scenario that allows for both training and assessment across the breadth and depth of the various performance elements we have identified. However, the availability of new approaches and technologies offers the potential for us to disentangle the training benefits and assessment of drill effectiveness, and to potentially augment EDs, in order to ensure more consistent, comprehensive (and, therefore, valuable) drill coverage. In the next Sections, we describe the methodology and main findings of the NEED project, which focused on (1) obtaining a wide-ranging picture of the current state of EDs (as delivered in practice), (2) a discussion of the limitations of current EDs, and (3) a comprehensive survey of alternative evacuation models in order to explore their strengths and weaknesses.

4. Community consultation methodology

The project was structured around a number of in-person workshops (held between late 2018 and early 2019, two in Canada and two in the UK), with the aim of addressing the core questions described in Section 2 [14]. Whilst acknowledging the difficulty of obtaining comprehensive coverage in terms of expertise and geographical location, we sought a range of international expertise that covered fire safety, human behavior, regulatory structures, training technologies and simulation tools, as well as training and learning analytics. A list of potential experts from academia, industry, regulatory organizations, and end users (e.g., airport authorities, city councils) was drawn up; ultimately, 30 experts from 20 different organizations and five countries participated in the workshops. These experts came from a range of different backgrounds: 36% of workshop participants were affiliated with universities, 30% with government bodies, 17% with industry, 10% with end users, and 7% with regulators. The full list of participants is supplied in [15], and we provide a copy at `https://doi.org/10.6084/m9.figshare.c.4876116.v1`

Given the variety of expertise, contributions addressed issues of both theory and practice. This was important, as (a) a stronger theoretical basis to our understanding of ED dynamics is critical to enhance practice, and (b) current practical limitations must be understood in order to both identify opportunities and to help to generate a stronger case for the need to improve the theoretical understanding of EDs.

The first two workshops set up the project infrastructure and established our three main areas of focus:

1. Training aspects of evacuation drills; how are they used to prepare individuals?
2. Assessment aspects of evacuation drills; how might we evaluate the performance of drills (and of individuals)?
3. Other ways of training and assessing; What are the alternative approaches to existing evacuation drills?

The third workshop formed the main core of the community consultation process (the final workshop mainly involved a presentation of our findings to relevant stakeholders and interested parties, with appropriate feedback taken for incorporation into the final version of the project roadmap).

Our enquiry was informed by a prior framework for the assessment of egress drills and possible alternative models/methodologies [14]. In particular, we were interested in the "considerations" listed in the paper, which may be interpreted as *evaluation criteria*. One of the key goals of the third NEED workshop was to obtain a set of evaluation criteria for both training and assessment methods, *without biasing the discussion* by presenting, *a priori*, the original criteria from [14]. We then assessed the closeness of fit between the original criteria and those we obtained from participants.

In order to explore these issues, during the third workshop, we asked participants to specifically consider the following main questions:

- (Q1) what are the training and assessment methods currently used, and what are their main positive and negative aspects?
- (Q2) What are the evaluation criteria for training and assessment, and how do we measure these criteria?

We also asked a number of supplementary questions:

- What are the most important criteria? Can we rank them in order of importance?
- What alternative methods for training and assessment might exist?

At the beginning of the third workshop, participants were introduced to the background of the project, agreed to sessions being (audio) recorded, gave informed consent for their input to be used, and were then divided into four mixed (in terms of background) breakout groups (each containing approximately four people) to discuss the questions listed above. At the beginning of the breakout session, each group was asked to assume that there are no restrictions on the number/type of methods available to train/assess occupants participating in EDs. Separate breakout sessions were held for training and assessment, in order to disentangle these two aspects (that is, each set of groups worked first on training, and then on assessment). Each group self-nominated a chair (with responsibility for keeping the discussion focussed on the main questions) and a rapporteur. In order to trial the data visualisation technique used in Section 6, Group 4 was asked to use the “radar diagram” method to record its findings.

For Q1, participants were first asked to describe their understanding of the current state-of-the-art in ED, and to describe positive and negative aspects. For Q2, participants were first primed with two example evaluation criteria (e.g., “cost”), and then were tasked to identify (and, where possible, rank according to their importance) *additional* potential criteria. Each breakout group received different primes in order to avoid biases in the criterion selection; these primes were either “Ethical and financial”, “Methodological and supervisory”, or “Statistical and pedagogical”; and one group did not receive any primes. The “Ethical prime recognised there may be a need to consider the risks associated with the drills, compared to the benefits; “Financial recognised that there may be various monetary costs associated with the drill; “Methodological recognised that there may be challenges in ensuring that accurate data can be collected at the required level of precision for a particular scenario; “Supervisory recognised the desire to independently monitor the outcomes of the approach (and the difficulty of doing so); “Statistical recognised the desire for the outcomes of the drill to have statistical validity, and “Pedagogical recognised the desire to ensure that attention can be focussed on all tasks at the desired level of scrutiny.

The priming example criteria were selected based on [14]. Participants then followed a structured discussion in which they addressed each of the questions. The results of these breakout sessions were recorded, summarized and reported back to the participants at the end of the workshop. A final discussion session was then used to agree on a consensus set of evaluation criteria, and to (where possible) rank them in order of importance.

The summary notes taken by each group are available at `https://doi.org/10.6084/m9.figshare.c.4876116.v1`

5. Consultation results

For reasons of space and clarity, we omit the full description of our results (which may be found in [15]). Here, we simply summarise the main findings concerning question Q1 (What training/assessment methods are currently used, and what are their pros and cons?), and (for Q2) describe the close alignment between the original set of evaluation criteria [14] and the criteria identified during our workshops. We first discuss the Q1 "state of the art" findings of the "training" session, followed by the findings of the "assessment" session, and then summarise the combined Q2 evaluation criteria discussions. Here, we emphasise that these are sometimes subjective opinions expressed by workshop participants, but we have tried, wherever possible, to represent only those themes and opinions that were sufficiently well-supported (either by evidence or consensus).

5.1. Current training methods, and their limitations

We partitioned existing methods into two categories: (1) "traditional drills, and (2) teaching-based approaches. While both categories are inherently artificial, the first is more experiential in nature, while the second tends to emphasise theory and pedagogy. All of these methods essentially explore "What if? scenarios, in order to establish the individual capacity to adapt and to make decisions under different conditions. However, there is the potential for different scenarios to be overlooked/excluded, or to be conflated. The performance of specific actions is related to individual roles within an evacuation procedure, and may be sensitive to specific factors related to the role, organization, procedure or occupancy. Existing methods are often applied at different levels (task-oriented, individual acts, procedural enactment) and to different groups [14]. They are often scheduled at different frequencies, depending on assessed risk, occupancy type, and the role(s) of the population under consideration.

EDs are often seen as the "standard method, with Fire Evacuation Officers (FEOs) assisting occupants to reach a place of safety. Participants reported on a combination of announced (warning provided ahead of time) and unannounced drills. Partial information is often provided, indicating that something will occur, but not precisely when. We may also see partial drills, based on zonal evacuation. Recording of drills provides opportunities to support subsequent debriefing and / or simulation studies, and may allow for comparison between individual actions and strategic objectives. EDs allow us to test whether or not processes are actionable. Drills also are used to provide knowledge of a building by forcing occupants to experience areas of the building that they were not previously aware of. We also see "FEO-only drills, involving only key personnel. Our workshop participants identified a number of limitations of traditional EDs with regard to training purposes, based on their personal and professional expertise and knowledge:

- Full-scale EDs pose serious safety challenges (e.g., the risk of crush injuries occurring in bottlenecks [23]). Methods are often selected based on risk of injury, rather than on the training benefits provided.

- Different occupancies have different population / procedures / staff/ risks, even within the same occupancy class. Concern was expressed regarding the physical/emotional consequences of including/excluding certain sub-populations (e.g., those with mobility impairments). Implications range from a perception of being patronised to degrading / uncomfortable / dangerous experiences.

- When releasing partial information ahead of a drill (e.g., to marshals), it is difficult to ensure that information does not propagate beyond a specified set of individuals.

- Loss of building functionality during a drill can incur direct overheads, and also incur reputational damage (if, for example, a large retail store is emptied of customers, the loss of revenue could be significant). Also, organizational/logistical costs may accrue in terms of loss of functionality and downtime, business continuity may be impacted, and an organization may suffer significant reputational damage.

- If drills last for longer than 20 minutes then participants may lose focus, and either mentally or physically disengage somewhat undermining the purpose of the event. (While the figure of 20 minutes was reported by a participant and may be challenged, the principle that prolonged events lead to a loss of interest seems a reasonable assumption.) There are also considerations of discomfort and impact on health (physical and emotional) this precludes consideration of a number of realistic scenarios.

A complementary widespread approach is based on more pedagogical methods, using presentations, videos, online training (e.g., 15 minute module, taken annually, with as many attempts as required for a pass). It may be the case that a relatively "low tech playback of a video (showing consequences of decisions, etc.) is sufficient to inform / practice decision making and provide feedback to participants. Such methods are generally seen as flexible and convenient, relatively low cost, and easy to use. They also allow for easy individual recording of completion of mandatory training (by keeping a register or online log of who has completed it). In addition to these methods (which are mainly concerned with imparting knowledge), it is also possible to generate relatively realistic time-induced pressure on decision making using table-top or simulation-based approaches. Our workshop participants identified the following limitations of teaching-based approaches (also discussed here [14]):

- Mostly knowledge-based they do not provide the physical experience of an actual evacuation, and may therefore be seen by participants as less "realistic. It is therefore unclear how well the training transfers to real situations in the physical world.
- Often based on solitary completion, and do not require any interaction with others. This may be significant in a real-world scenario, as interpersonal interactions will play a role.

- Frequently viewed by participants as an unmotivating "box ticking exercise, which may diminish engagement and therefore the impact of training.

- May not require consistent and meaningful engagement with the material, and may be possible for participants to "pass without active learning. In other words, it can be unclear how much information trainees are retaining and applying to their own personal situation.

5.2. Current assessment methods, and their limitations

According to feedback from workshop participants, the most common method for assessing procedures is the traditional ED, and this often focuses on final outcomes (i.e., the final evacuation time - were occupants evacuated from the building safely and within a specified time limit?), without considering the internal dynamics of the exercise. However, analysis of exercises does sometimes consider qualitative aspects (such as "orderliness), as well as other quantitative aspects such as pedestrian flow and/or exit usage, speed of movement, etc.

EDs are useful in that they require participants to physically enact procedures. This has two main benefits: (1) they create observable behaviour (generating usable data), and (2) they are capable of generating uncertainty/randomness that may not be captured or encoded in simulations or non-embodied methods. They also have the potential for the greatest possible level of ecological validity (for a range of scenarios) [14, 24].

Other methods that are used (usually in specialist applications or situations of particular concern) include computational simulation (often in conjunction with a traditional drill) [10], or micro-level assessment of specific components of evacuation (such as the evacuation of people with mobility impairments), as well as treatment of hypothetical situations using mechanisms such as tabletop exercises, walk-throughs, laboratory experiments, and virtual/augmented reality experiments. When a traditional ED is used, a variety of data capture methods may be employed to record information about its execution. These include CCTV/video capture [25], head counts, marshal feedback, regulatory body observers, surveys (e.g., questionnaires [26]), post-event interviews and debriefing [27] (for a comprehensive survey of data collection methods, see [28]). The data collected varies according to the method employed.

The workshop participants identified a number of specific limitations of traditional EDs with regard to assessment:

- Existing methods for assessing evacuation procedures can be very outcome-driven, in that they tend to treat the activity as a “black box. If undue focus is placed on “bottom line metrics (and consideration is not made of “internal” features of the activity, such as congestion points, etc.), then this may lead to an implicit / explicit bias in the design of the exercise.

- There is always the issue that a drill may fulfil (sometimes by accident) a specific criterion without participants necessarily adhering to procedures. For example, occupants might safely reach assembly points in time but not follow the shortest or safest egress route. In essence, it would then not be clear whether the outcome (successful or otherwise) was achieved through the use of the intended procedure, or simply coincidentally, given other non-procedural factors.

- Drills do not interrogate all procedures, operational environments or scenarios, because of the exclusion of certain populations or execution in favourable conditions, announced versus unannounced status, and so on, so data collection is inherently partial. We must also consider the extreme cost of examining all possible factors that might affect performance.

- Because of the nature of drills, it is very difficult to establish longitudinal reliability, as they are effectively “one off events. Indeed, the authors are unaware of any documented attempts to do so.

5.3. Evaluation criteria for both training and assessment

When we summarised the group discussions, we observed a very close match between the criteria we generated during the workshops and the original criteria in [14]. In Table 1 we show the alignment between the original criteria and those subsequently obtained through our workshop discussions. Of course, more criteria than the ones identified here are conceivable and some might only apply to training aspects and some only to assessment. For an expanded discussion of the criteria, including examples of questions that they might provoke, we refer the reader to the NEED roadmap [15].

For training criteria, during the third workshop we were able to rank the criteria in descending order of perceived significance (so "Adherence to regulatory requirements" is the most important), according to the aggregated view of our participants. The ranking was generated by consensus in individual break-out groups, and then agreed in a whole-group debriefing session. We emphasise that this is only an approximate ranking, and that the specific ordering of criteria will differ according to the organization, building type, population(s), etc. For example, if we consider a nuclear power plant, then risk (to health and safety in case of emergencies), will be more highly-weighted than cost (we would hope), whereas a stock exchange building might carry relatively low risk but incur hugely significant financial costs if impacted by a drill. For assessment, we found it more difficult to agree on a definitive consensus rank ordering, so we do not number these.

Given that the set of criteria obtained via workshop criteria closely (and independently) match those of [14], we conclude that we have generated a robust set of criteria against which both training and assessment methods may be assessed. For clarity, in what follows we use the original set of criteria from [14], noting that these serve as "shorthand" definitions that may be expanded upon by incorporating the criteria (and their descriptions) that we subsequently generated.

6. Criteria-based framework for comparing approaches

Several alternatives to traditional EDs for both training and assessment have been identified [14], including simulations, laboratory experiments, immersive (Virtual/Augmented Reality) technologies, table-top exercises, mental

Original criteria [14]	**Our training criteria**	**Our assessment criteria**
Financial/organizational.	Cost / opportunity cost (4). Convenience / accessibility (7). Performance / reliability / availability of procedural infrastructure (8).	Cost.
Ethical.	Risk (physical / emotional/ dignity), safety, and ethical concerns (5).	Safety.
Perceived credibility.		User satisfaction with procedure.
Scope.	Scenario coverage / transferability / adaptability (3). Reach / inclusion of populations (6).	Robustness, adaptability and scope of procedures.
Potential insights.	Data collection and subsequent actions (9).	Consistency, accuracy and predictability of results Actual versus expected performance.
Third party scrutiny.	Adherence to regulatory requirements (1).	Adherence to protocol.
Statistical.	Data collection and subsequent actions (9).	Data collection.
Pedagogical.	Achievability of training objective(s) (2).	

Table 1: Alignment between Gwynne, *et al.* [14] criteria and those identified in our workshop.

rehearsals, walkaround sessions, briefings and scripted exercises. Each of these approaches has their relative merits, so here we demonstrate how they may be easily compared with one another, using our list of criteria combined with workshop participant responses to Q4, concerning alternative approaches. In order to illustrate our approach, we compare the following alternatives to the drill model: (1) Simulation, (2) Immersive technology (virtual and augmented reality), and (3) and Table-top approaches (for full results of all approaches discussed please see [15]).

In order to systematically visualize the comparisons, each method was rated by the workshop participants on an arbitrary 7-point Likert scale ranging from 0 (“low”) to 3 (“high”) in 0.5 increments. Alignment with points on each scale was assessed against individual criteria; so, for example, for the “Ethical” criterion, a “low” level of alignment meant that a method had potential negative ethical issues, whereas a “high” level of alignment meant that the method was rated by participants as ethically sound. Similarly, for “Potential insights”, a low level of alignment meant that participants believed that a method had a relatively low potential for generating useful insights, and so on. The alignment across criteria was then plotted in radar diagrams, with each spoke denoting an evaluation criterion from the set (Financial, Ethical, Credibility, Scope, Insights, Scrutiny, Statistical, Pedagogical); see previous Section. Note that the order in which the criteria are presented is completely arbitrary, and the weight of each criterion may vary across circumstances. Further, the directions of the scales are not uniform across criteria (e.g., for the ”financial” criterion, higher costs are associated with a lower score; for the ”credibility” criterion, higher credibility indicates better alignment). In addition, the ratings themselves represent consensus expert opinions, and have similar limitations as other forms of subjective assessments. Despite these limitations, the visualization provides a possible framework for a systematic comparison of different methods. We now give three examples of such comparisons, based on assessments obtained from participants in our workshop.

6.1. Drill versus simulation

Computational simulation tools are frequently used as part of a performance-based assessment to quantify evacuation performance (along with prediction of the fire conditions) [14, 29]. They are used to calculate the time for populations to reach a place of safety, the performance elements that contributed to this (e.g. time to travel, congestion levels, etc.) and the conditions experienced. They are therefore becoming more familiar tools to fire safety practitioners and also shaping expectations as to the refinement of results and the range of scenarios that might be examined [9, 30, 31]. The primary application of simulation tools is to calculate estimates for the required and available safe egress times [32]; however, they are increasingly being used to provide benchmarks as to potential evacuation times given the use of different procedures and different scenarios [33, 34]. Such output, provided in a suitable format might support training efforts and may also provide a benchmark against which evacuation performance might be compared. Figure 1 compares the perceived performance of simulation tools with evacuation drills.

1. Financial: EDs were estimated by participants as having a smaller financial burden than simulation tools. We surmise that this assessment addresses the labour cost involved in configuring, executing and analysing the results produced [14]. It may also include the purchase of the simulation tool (if it is not freely available), should the analysis be completed in-house. In comparison, the resources involved in completing an ED would likely be relatively small, if the loss of functionality of the building (and associated services) during the drill is not included (however, as we have noted earlier, these costs may be significant, so it may not be realistic to discount them).
2. Ethical: EDs were rated by participants as having a lower ethical rating; i.e. that they pose more significant concerns for safety and well-being. This appears logical, given that the simulation tools do not involve active participants and are not necessarily shared with human subjects [13]. As

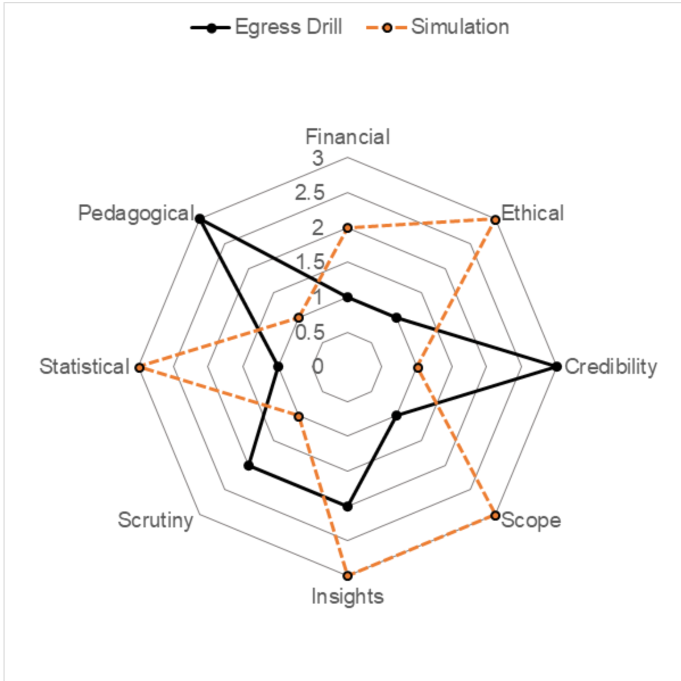

Figure 1: Radar diagram plotting alignment of EDs and simulation tools with eight evaluation criteria. Data are presented on an arbitrary 7-point Likert scale ranging from 0 (no alignment with criterion) to 3 (perfect alignment with criterion).

scenarios become more realistic and include more representative conditions (e.g. route loss, a sense of urgency, etc.), so the risks posed to the participants would increase.

3. Credibility: The experts clearly rated EDs as having far superior credibility levels in comparison to simulation tools. One potential explanation for this is user familiarity with EDs (and relative lack of familiarity with simulation tools)[1]. Another potential explanation could be the fact that the ED "model" of evacuations includes actual human participants (and associated decision-making), as opposed to computer-based agents (and

[1]Many simulation tools are based on data collected from drills and so it is suggested that by proxy, simulation tools risk importing some of the issues raised.

associated design decision-making) in simulation tools. The fact that the drill is only a model (an approximation) of a real event is often not considered [2], nor is the fact that simulations, in turn, are often developed and calibrated using drill data, thus embedding ED-specific uncertainty into the modelling method.

4. Scope: Experts assessed the simulation tool to be capable of representing a wider range of scenarios than an ED. This is likely due both to experience with the limited set of scenarios typically represented by EDs (e.g. all exits available, etc.) and the understanding that simulation tools rate high ethically (given the lack of risk they pose to participants) enabling them to subject the simulated evacuees to an array of conditions not accessible in real-life situations [14].
5. Insights: Experts rated simulation tools as being better able to provide insights - this is likely due to the fact that simulation tools might directly provide insights at the agent-level and summary aggregate levels, and that the output from the simulation is a direct reflection of the conditions experienced by the agents and the agent actions. In an ED, such metrics and relationships often have to be inferred and interpreted from expert observation or through surveys which involves some degree of participant reporting [14].
6. Scrutiny: Simulation tools were considered by participants to be less open to scrutiny by third party observers, compared to an ED [13]. Some experts pointed out that this was particularly the case for commercial software products, but less relevant to open-source tools. EDs can certainly be directly observed, allowing adherence to regulatory requirements to be established. This likely influenced this rating. However, what is less clear is how frequently such opportunities are taken; i.e. how frequently drills are independently observed.
7. Statistical: Our experts rated simulation tools as being able to conduct repeat trials more frequently than an ED approach. This implies that simulations can be repeated, producing more confidence in the results pro-

duced. EDs (even if deemed less expensive, as noted above) cannot easily be repeated given disruption to the building and due to the population learning from previous drill events [13].

8. Pedagogical: Experts clearly identified EDs as having a higher instructive and training potential. EDs involve human subjects (typically occupants of the structure in question) who experience the events of the ED directly. This allows them to be cognizant of the drill, their actions in and the outcomes. Simulation tools do not enable such experience - except through observing the results of the simulation to observe the implications of specific actions and procedures [13]. In addition, they provide indirect means of training or assessment; i.e. the output from the simulation might be used to demonstrate the consequences of certain actions / performance-levels enhancing training, or might provide benchmark performance levels against which the results of a drill might be compared.

6.2. Drill versus immersive technology

Emerging immersive technologies such as virtual reality (VR) and even more recently augmented reality (AR) have become increasingly popular for training and assessment tools (see, for example, [24, 35]). Two expectations are driving this trend: first, VR and AR promise effective, flexible and affordable training platforms. Second, VR and AR balance ecological validity and experimental control [7]. Much has been written about VR (for recent publications, see for instance, [36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48]). AR, however, has only recently been identified as a training and research tool [49, 50, 51]. Here, our comparison focusses primarily on VR using Head Mounted Display (HMD) technology, where users wear trackable headsets which can display virtual content in a first person perspective.

How do immersive technologies compare to traditional EDs? Figure 2 compares immersive technologies to EDs. We now describe, in detail, the rationale for our assessments against each criterion, which were obtained during NEED workshop discussions.

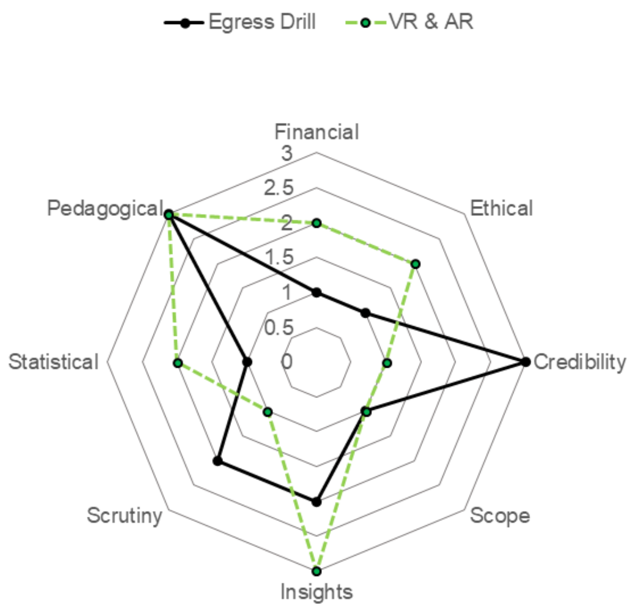

Figure 2: Radar diagram plotting alignment of EDs and VR and AR training tools. Data are plotted in the same way as in Figure 1.

1. Financial: EDs were estimated to have a lower financial burden than VR and AR training tools. Resources need to be dedicated to developing and evaluating the immersive training content, as well as updating it to changes in procedures. However, VR training tools become cost-effective when they are deployed *at scale*. Then the promise is that users can complete the training in small groups or individually, without disturbing the productivity of a building/organization as a whole.
2. Ethical: VR training was rated as having fewer ethical limitations compared to EDs. Traditional drills involve risks for trainees and often exclude vulnerable populations (e.g., mobility impaired occupants). However, VR has potential side effects (e.g., simulator sickness), and some users may not be able to participate (e.g., visually impaired trainees).

3. Credibility: The experts clearly rated EDs as having superior credibility over immersive technology. EDs have been established, and, despite their limitations, (see discussion for example in [14]), they are still considered to be the "gold standard" for training and assessing occupants. Alternative approaches need to show that they allow comparable assessments as drills, and that training translates from the virtual into the real world. Although some progress has been made on the former issue [8, 52, 53], and significant evidence has been found for the latter [54, 55], future research is clearly necessary in this area.
4. Scope was identified as a relative weakness of both methods, the general limitation being that the scope of a specific drill or VR training might be too narrow to prepare occupants for or assess occupants over a wide range of possible scenarios.
5. Insights: VR training has greater potential to generate insights compared to EDs. EDs can provide valuable insights, given accurate observation and documentation. Typically, VR training can provide more granular insights (e.g., at the level of the individual trainees as opposed to more global observations, such as the total evacuation time of a building).
6. Scrutiny: This was seen as a relative weakness of VR training compared to EDs. In theory, adherence to regulatory requirements and protocols is directly observable in EDs, whereas this is not possible in VR.
7. Statistical: VR studies allow for more systematic, fine grained and automatized data collection methods compare to drills. However, future developments in automatized data collection methods may facilitate improved data quality in traditional drills.
8. Pedagogical: Both methods strengths lies in their pedagogical value. Performance before and after training are easily observable, and feedback can be provided with relative ease to trainees (given accurate observation).

6.3. Drill versus table-top exercises

Tabletop exercises refer to forms of training in which participants collaboratively work through hypothetical scenarios in a classroom setting and assume and discuss roles and tasks [56, 57]. Often, a facilitator may guide the session, assign roles, and scenario development.

How are table-top exercises evaluated compared to EDs? 3 compares table-top exercises to EDs. Below, we describe, in detail, the rationale for our assessments against each criterion, which were obtained during NEED workshop discussions.

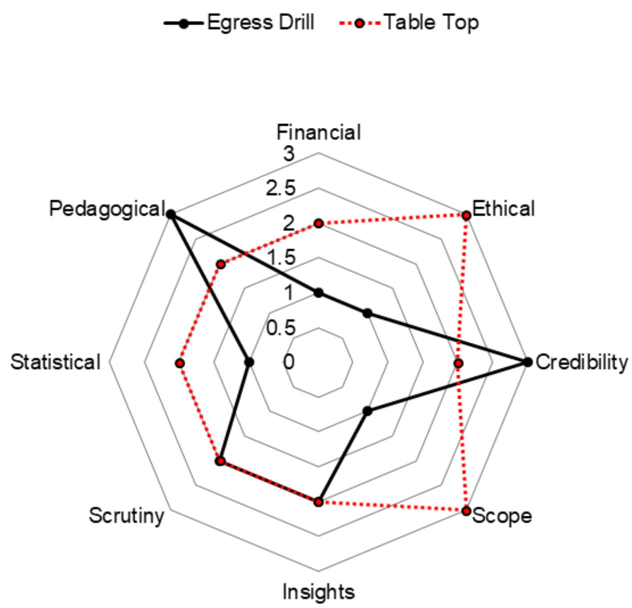

Figure 3: Radar diagram plotting alignment of EDs and tabletop exercises. Data are plotted in the same way as in Figure 1.

1. Financial: Participants rated EDs to be better aligned with this criterion than tabletop exercises. For instance, costs for tabletop exercises may

occur for planning and facilitation. However, they typically require fewer staff involvement and less building downtime.

2. Ethical: Similar to immersive technologies, tabletop exercises were rated as having fewer ethical limitations compared to EDs.
3. Credibility: Tabletop exercises were seen as less credible than EDs for similar reasons as immersive technologies (less realism and not a true test of operational capability).
4. Scope: The ability to clearly define and dynamically adjust scope was rated as a strength of tabletop exercises in comparison to EDs.
5. Insights: Tabletop exercises and EDs were both rated by participants to be moderately aligned with this criterion. However, the type of insights that may emerge from the two types of training might differ. In tabletop exercises, insights can be gained, for example, regarding the decision-making processes in teams.
6. Scrutiny: Both forms of training appear to be moderately aligned with this criterion; Although some insights can be gained into, for instance, procedural knowledge, tabletop exercises do not allow for a true test of adherence to protocols.
7. Statistical: The controlled setting of tabletop exercises allow for relatively easy data collection compared to EDs. However, this data is typically not behavioral and consequently limited in ecological validity.
8. Pedagogical: While this is a clear strength of EDs, the participants found tabletop exercises to be only moderately aligned with this criterion.

7. Conclusions and Recommendations

We conclude by summarising the main problems with traditional EDs that we have identified, and propose one possible way forward that might begin to address these. EDs are pre-planned simulations of emergency evacuations for specific scenarios, with the aim of improving and assessing the performance of individuals involved. Although drills are informed by safety legislation and code, their merits are still not well-understood, and their impact on evacuation performance is not well-characterized [14]. The benefits of such drills are unclear, given that:

1. Drills carry inherent *risks* to participants and significant costs (e.g., temporary loss of building functionality, added liability).
2. They are *inconsistently performed* and not fully exploited to meet their twin objectives of training and performance assessment. For example, subpopulations are often not included in drills such as mobility impaired occupants, or mission-critical staff, affecting the potential for realistic training / assessment.
3. The nature of potential emergency scenarios is *constantly evolving* (e.g., climate change extreme events, terrorism, active shooter incidents), as are population demographics [33].
4. Data is usually collected manually and often based on subjective assessments (e.g., manual timing; qualitative performance ratings)

Nonetheless, EDs are still seen as key components of safety planning / building certification. Given this reliance, it is vital to understand whether EDs, as currently performed, are fit for purpose, and if they need to be enhanced (or even replaced) in a cost-effective manner. The implication on practice is that emerging immersive / simulation technologies offer potential alternatives to the existing drill model, and could potentially mitigate the challenges mentioned above. In terms of future research focus, the effectiveness, credibility and validity of these new approaches needs to be assessed against the existing approach. The findings from our NEED workshops have highlighted the limitations of the current approach and the significant opportunities offered by emerging technologies to improve the safety of building occupants, while reducing the negative side-effects of drills. A particular strength of our approach is that we gathered multidisciplinary and multinational perspectives, by including experts from five countries and different types of organisations (safety practitioners, regulatory bodies, academia, etc.) We also identified a number of criteria against which different methodologies might be assessed (e.g., cost, realism, or credibility). We propose that future work should include the following objectives:

1. Establish an evidence-based methodology (case-control approach) for assessing evacuation drills and alternatives; i.e. to measure the effectiveness of different approaches in a methodical manner.
2. Harness novel objective and automatable approaches to data capture and analytics to better characterize performance (using smart sensors, artificial intelligence, computer vision, and machine learning).
3. Develop alternatives to the current drill model, based on emerging immersive / simulation technologies (e.g., virtual / augmented reality) that specifically target the limitations of traditional EDs, and compare its performance to the status quo.
4. Develop guidance for regulatory bodies on the application and cost-benefits of each approach (e.g., relative performance gain, loss of individual / building time) for different scenarios.

It should be noted that these suggestions would make use of existing understanding and technologies, primarily requiring the integration of such approaches, rather than new developments. The primary challenges will then be in updating regulations and guidance such that these approaches are adopted, educating potential users to ensure they are suitably applied, and managing their effectiveness once employed.

Given the importance of EDs to life safety, it is critical that we better understand their effectiveness and identify alternative approaches if limitations cannot be addressed in practice. The future work highlighted above, suggested by our analysis of the compiled feedback of workshop participants, will go some way towards achieving these goals.

Acknowledgements

We acknowledge support from the National Research Council of Canada (NRC) and the UK Arts and Humanities Research Council (AHRC) for the

"Networking activity for Enhanced Evacuation Drills (NEED)" project, contract number A1-015092 (NRC), project reference AH/S012737/1 (AHRC). CNvdW is supported by the European Union Horizon 2020 research and innovation programme: Marie Sklodowska-Curie grant agreement No 748647. ER's egress work is supported by FORMAS (Swedish Research Council for sustainable development) through the "Building Egressibility in an Ageing Society" project. We also thank all of the participants in the NEED workshop series for their invaluable contributions to the project.

References

[1] R. Peacock, J. Averill, E. Kuligowski, Stairwell evacuation from buildings: what we know we dont know, in: Pedestrian and Evacuation Dynamics 2008, Springer, 2010, pp. 55–66.

[2] R. D. Peacock, B. L. Hoskins, E. D. Kuligowski, Overall and local movement speeds during fire drill evacuations in buildings up to 31 stories, Safety Science 50 (8) (2012) 1655–1664.

[3] G. Proulx, I. M. Reid, Occupant behavior and evacuation during the Chicago Cook County Administration Building fire, Journal of Fire Protection Engineering 16 (4) (2006) 283–309.

[4] T. Shields, K. Boyce, A study of evacuation from large retail stores, Fire Safety Journal 35 (1) (2000) 25–49.

[5] T. Shields, K. Boyce, N. McConnell, The behaviour and evacuation experiences of WTC 9/11 evacuees with self-designated mobility impairments, Fire Safety Journal 44 (6) (2009) 881–893.

[6] Z. Feng, V. González, R. Amor, R. Lovreglio, G. Cabrera-Guerrero, Immersive virtual reality serious games for evacuation training and research: A systematic literature review, Computers & Education 127 (2018) 252–266.

[7] M. Kinateder, E. Ronchi, D. Nilsson, M. Kobes, M. Müller, P. Pauli, A. Mühlberger, Virtual reality for fire evacuation research, in: 2014 Federated Conference on Computer Science and Information Systems, IEEE, 2014, pp. 313–321.

[8] M. Kinateder, W. Warren, Social influence on evacuation behavior in real and virtual environments, Frontiers in Robotics and AI 3 (2016) 43.

[9] S. Gwynne, E. R. Galea, M. Owen, P. J. Lawrence, L. Filippidis, A review of the methodologies used in the computer simulation of evacuation from the built environment, Building and Environment 34 (6) (1999) 741–749.

[10] I. von Sivers, A. Templeton, F. Künzner, G. Köster, J. Drury, A. Philippides, T. Neckel, H.-J. Bungartz, Modelling social identification and helping in evacuation simulation, Safety Science 89 (2016) 288–300.

[11] M.-Y. Cheng, K.-C. Chiu, Y.-M. Hsieh, I.-T. Yang, J.-S. Chou, Y.-W. Wu, BIM integrated smart monitoring technique for building fire prevention and disaster relief, Automation in Construction 84 (2017) 14–30.

[12] K. Wang, X. Shi, A. P. X. Goh, S. Qian, A machine learning based study on pedestrian movement dynamics under emergency evacuation, Fire Safety Journal 106 (2019) 163–176.

[13] S. Gwynne, K. Boyce, E. Kuligowski, D. Nilsson, A. Robbins, R. Lovreglio, Pros and cons of egress drills, Interflam 2016.

[14] S. Gwynne, E. Kuligowski, K. Boyce, D. Nilsson, A. Robbins, R. Lovreglio, J. Thomas, A. Roy-Poirier, Enhancing egress drills: Preparation and assessment of evacuee performance, Fire and Materials.

[15] M. Amos, N. Benichou, S. Gwynne, M. Kinateder, NEED: Networking Activities for Enhanced Evacuation Drills - Roadmap for Enhanced Evacuation Drills, National Research Council of Canada Report A1-015092.1. doi: 10.4224/40001242. (2019).

[16] L. Catovic, C. Alniemi, E. Ronchi, A survey on the factors affecting horizontal assisted evacuation in hospitals, Journal of Physics: Conference Series 1107 (7) (2018) 072001.

[17] A. Rigos, E. Mohlin, E. Ronchi, The cry wolf effect in evacuation: A game-theoretic approach, Physica A: Statistical Mechanics and its Applications 526 (2019) 120890.

[18] M. Ramirez, K. Kubicek, C. Peek-Asa, M. Wong, Accountability and assessment of emergency drill performance at schools, Family & Community Health 32 (2) (2009) 105–114.

[19] V. Alonso-Gutierrez, E. Ronchi, The simulation of assisted evacuation in hospitals, in: Fire and Evacuation Modelling Technical Conference, Malaga, Spain, 2016.

[20] S. Breznitz, Cry wolf: The Psychology of False Alarms, Psychology Press, 2013.

[21] J. M. Berlin, E. D. Carlström, Collaboration exercises: What do they contribute? –a study of learning and usefulness, Journal of Contingencies and Crisis Management 23 (1) (2015) 11–23.

[22] Guidelines to developing emergency action plans for all-hazard emergencies in high-rise office buildings, National Fire Protection Association (2014).

[23] C. Johnson, Lessons from the evacuation of the World Trade Centre, 9/11 2001 for the development of computer-based simulations, Cognition, Technology & Work 7 (4) (2005) 214–240.

[24] R. Lovreglio, V. Gonzalez, R. Amor, M. Spearpoint, J. Thomas, M. Trotter, R. Sacks, The need for enhancing earthquake evacuee safety by using virtual reality serious games, in: Lean & Computing in Construction Congress, 2017, pp. 1–9.

[25] H. Najmanová, E. Ronchi, An experimental data-set on pre-school children evacuation, Fire Technology 53 (4) (2017) 1509–1533.

[26] L. Chen, T.-Q. Tang, H.-J. Huang, Z. Song, Elementary students evacuation route choice in a classroom: A questionnaire-based method, Physica A: Statistical Mechanics and its Applications 492 (2018) 1066–1074.

[27] R. J. R. Salway, Z. Adler, T. Williams, F. Nwoke, P. Roblin, B. Arquilla, The challenges of a vertical evacuation drill, Prehospital and Disaster Medicine 34 (1) (2019) 25–29.

[28] T. Rinne, K. Tillander, P. Gronberg, Data collection and analysis of evacuation situations, VTT Technical Research Centre of Finland, Research Notes 2562 (2010).

[29] E. Ronchi, E. D. Kuligowski, D. Nilsson, R. D. Peacock, P. A. Reneke, Assessing the verification and validation of building fire evacuation models, Fire Technology 52 (1) (2016) 197–219.

[30] E. D. Kuligowski, Computer evacuation models for buildings, in: SFPE Handbook of Fire Protection Engineering, Springer, 2016, pp. 2152–2180.

[31] E. D. Kuligowski, R. D. Peacock, P. A. Reneke, E. Wiess, C. R. Hagwood, K. J. Overholt, R. P. Elkin, J. D. Averill, E. Ronchi, B. L. Hoskins, et al., Movement on stairs during building evacuations, NIST Technical Note (1839) 122.

[32] D. Purser, ASET and RSET: addressing some issues in relation to occupant behaviour and tenability, Fire Safety Science 7 (2003) 91–102.

[33] E. Ronchi, E. Kuligowski, S. Gwynne, Guest editorial: Special issue on advances in evacuation modelling, Fire Technology 55 (2019) 387389.

[34] E. D. Kuligowski, S. M. Gwynne, What a user should know when selecting an evacuation model, NIST Technical Report 101157 (2005).

[35] F. M. Williams-Bell, B. Kapralos, A. Hogue, B. Murphy, E. Weckman, Using serious games and virtual simulation for training in the fire service: a review, Fire Technology 51 (3) (2015) 553–584.

[36] W. Ahmad, A. Sarlan, N. Rauf, Escape: Interactive fire simulation and training for children using virtual reality, in: International Conference of Reliable Information and Communication Technology, 2018, pp. 670–679.

[37] E. Bourhim, A. Cherkaoui, Selection of optimal game engine by using AHP approach for virtual reality fire safety training, in: International Conference on Intelligent Systems Design and Applications, 2018, pp. 955–966.

[38] Ü. Çakiroğlu, S. Gökoğlu, Development of fire safety behavioral skills via virtual reality, Computers & Education 133 (2019) 56–68.

[39] R. Clifford, S. Jung, S. Hoerrnann, M. Billinqhurst, R. Lindeman, Creating a stressful decision making environment for aerial firefighter training in virtual reality, in: 26th IEEE Conference on Virtual Reality and 3D User Interfaces, 2019, pp. 181–189.

[40] S. Danial, J. Smith, F. Khan, B. Veitch, Human-like sequential learning of escape routes for virtual reality agents, Fire Technology 55 (3) (2019) 1057–1083. doi:10.1007/s10694-019-00819-7.

[41] Z. Liu, Q. Zhao, L. Sun, Y. Luo, Application of virtual human factor verification to fire accidents at the main control room in nuclear power plant, in: International Conference on Man-Machine-Environment System Engineering, Springer, 2019, pp. 885–892.

[42] S. Nahavandi, L. Wei, J. Mullins, M. Fielding, S. Deshpande, M. Watson, S. Korany, D. Nahavandi, I. Hettiarachchi, Z. Najdovski, et al., Haptically-enabled VR-based immersive fire fighting training simulator, in: Intelligent Computing - Proceedings of the Computing Conference, Springer, 2019, pp. 11–21.

[43] T. Nilsson, T. Roper, E. Shaw, G. Lawson, S. Cobb, H. Meng-Ko, D. Miller, J. Khan, Multisensory virtual environment for fire evacuation training, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, 2019, p. INT044.

[44] S. Ooi, T. Tanimoto, M. Sano, Virtual reality fire disaster training system for improving disaster awareness, in: Proceedings of the 2019 8th International Conference on Educational and Information Technology, ACM, 2019, pp. 301–307.

[45] K. L. Rossler, G. Sankaranarayanan, A. Duvall, Acquisition of fire safety knowledge and skills with virtual reality simulation, Nurse Educator 44 (2) (2019) 88–92.

[46] U. Rüppel, K. Schatz, Bim-based virtual training environment for fire-fighters, in: Proceedings of the 13th International Conference on Computing in Civil and Building Engineering & 2008 International Conference on Information Technology in Construction, 2010, p. 23.

[47] E. Shaw, T. Roper, T. Nilsson, G. Lawson, S. V. Cobb, D. Miller, The heat is on: exploring user behaviour in a multisensory virtual environment for fire evacuation, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, 2019, p. 626.

[48] S. A. Fathima, S. J., R. Jenice Aroma, Simulation of fire safety training environment using immersive virtual reality, International Journal of Recent Technology and Engineering 7 (4) (2019) 347–350.

[49] H. Engelbrecht, R. Lindeman, S. Hoermann, A swot analysis of the field of virtual reality for firefighter training, Front. Robot. AI 6: 101.

[50] J. Potts, T. Sookdeo, J. Westerheide, D. Sharber, et al., Enhanced augmented/mixed reality and process safety applications, in: SPE Middle East Oil and Gas Show and Conference, Society of Petroleum Engineers, 2019.

[51] X. Li, W. Yi, H.-L. Chi, X. Wang, A. P. Chan, A critical review of virtual and augmented reality (VR/AR) applications in construction safety, Automation in Construction 86 (2018) 150–162.

[52] M. Moussaïd, M. Kapadia, T. Thrash, R. Sumner, M. Gross, D. Helbing, C. Hölscher, Crowd behaviour during high-stress evacuations in an immersive virtual environment, Journal of The Royal Society Interface 13 (122) (2016) 20160414.

[53] T. Thrash, M. Kapadia, M. Moussaid, C. Wilhelm, D. Helbing, R. W. Sumner, C. Hölscher, Evaluation of control interfaces for desktop virtual environments, Presence: Teleoperators and Virtual Environments 24 (4) (2015) 322–334.

[54] Y. Gao, V. Gonzalez, T. Yiu, The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review, Computers and Education 138 (2019) 101–115. doi:10.1016/j.compedu.2019.05.003.

[55] M. Kinateder, P. Pauli, M. Müller, J. Krieger, F. Heimbecher, I. Rönnau, U. Bergerhausen, G. Vollmann, P. Vogt, A. Mühlberger, Human behaviour in severe tunnel accidents: Effects of information and behavioural training, Transportation Research Part F: Traffic Psychology and Behaviour 17 (2013) 20–32.

[56] S. Renner, et al., Emergency exercise and training techniques, Australian Journal of Emergency Management 16 (2) (2001) 26.

[57] L. G. Holloway, Emergency preparedness: Tabletop exercise improves readiness, Professional Safety 52 (8) (2007) 48.