Evaluation of emergency preparedness exercises: the design of a questionnaire to measure staff perceptions

Elena A. Skryabina* and Naomi Betts

Behavioural Science Team,
Emergency Response Department Science and Technology,
Public Health England,
Porton Down, Salisbury,
Wiltshire, SP4 0JG, UK
Email: elena.skryabina@phe.gov.uk
Email: naomi.betts@phe.gov.uk
*Corresponding author

Gabriel Reedy

Faculty of Life Sciences and Medicine,
King's College London,
Waterloo Bridge Wing 5.14,
London, SE1 8WA, UK
Email: gabriel.reedy@kcl.ac.uk

Paul Riley

Emergency Preparedness and Response Support,
European Centre for Disease Prevention and Control (ECDC),
Gustav III: s boulevard 40,
169 73 Solna, Sweden
Fax + 46 (0)8 58 60 10 01
Website: www.ecdc.europa.eu
Email: Paul.Riley@ecdc.europa.eu

Richard Amlôt

Behavioural Science Team,
Emergency Response Department Science and Technology,
Public Health England,
Porton Down, Salisbury,
Wiltshire, SP4 0JG, UK
Email: richard.amlot@phe.gov.uk

Abstract: Evaluation is an essential part of health emergency preparedness exercises (HEPE) that allows identification of limitations in performance. Addressing limitations enhances preparedness. However, there is a lack of reliable and validated tools to assist with exercise evaluation. This study reports

the design and validation of a questionnaire to collect data from participants to study their experiences with HEPE and perceptions of their own and their organisation's emergency preparedness. Questionnaire test-retest reliability using ICC was checked ($N = 27$). Internal reliability using Cronbach's alpha is reported separately for discussion-based ($N = 97$) and operation-based exercises ($N = 238$). Analysis checked for discriminant validity and sensitivity to change. The questionnaire consists of four scales (parts): pre-exercise assessment, participant's perceptions, exercise feedback, and satisfaction with the exercise. All scales demonstrated good internal consistency for both exercise types (Cronbach's alpha: 0.672–0.940), but mixed test-retest agreement for the pre-exercise and exercise feedback scales. The questionnaire offers a valid and reliable tool for assessing healthcare staff perceptions of emergency preparedness and exercise satisfaction.

Keywords: HEPE; health emergency preparedness exercises; evaluation; questionnaire; survey; TTX; tabletop; drill; healthcare emergency responder; emergency preparedness; operation exercise; simulation exercise; healthcare emergency; mass casualty; major incident; disaster.

Reference to this paper should be made as follows: Skryabina, E.A., Betts, N., Reedy, G., Riley, P. and Amlôt, R. (2020) 'Evaluation of emergency preparedness exercises: the design of a questionnaire to measure staff perceptions', *Int. J. Emergency Management*, Vol. 16, No. 4, pp.317–346.

Biographical notes: Elena A. Skryabina, PhD, is a Research Fellow at Public Health England (PHE) Emergency Response Department (ERD). She leads on a number of research projects funded by the National Institute for Health Research (NIHR) for Health Protection Research Units (HPRU) in Emergency Preparedness and Response. The overaching aim of research is to enhance emergency preparedness through improved exercises and training. She works in close collaboration with Emergency Exercises and Training Teams at PHE ERD; the team delivers health-led exercises across the UK and Europe.

Naomi Betts, MSc, is a PhD student at King's College London working with Public Health England (PHE) Emergency Response Department (ERD) to research ways of enhancing discussion-based tabletop (TTX) emergency preparedness exercises. Specific focus of Naomi's research is to understand factors that affect participants learning from TTX to contribute to exercise design, delivery and evaluation. A longitudinal study will explore exercises immediate (post-exercise), medium term (6 months) and longer-term (12 months) impact on participants learning and emergency preparedness.

Gabriel Reedy, PhD is a Reader in Medical Education and the Educational Research Lead for the Simulation and Interactive Learning (SAIL) Centre, King's College London. Gabriel contributes to the Centre's educational programs, leads on the educational component of the Centre's research projects and helps to direct the educational and research works of the centre's medical education fellows.

Paul Riley, PhD is a Principle Expert Public Health Emergency Preparedness at ECDC. Paul specialises in developing processes to provide expert scientific support to first responders in emergencies and managing programs for emergency preparedness development and delivery for the European Commission. He is a Member of the British Standards Institute exercising and testing committee.

Richard Amlôt, PhD, Professor is a Programme Leader of the Behavioural Science Team in the Emergency Response Department at PHE. He leads a program of research that explores operational, behavioural and psychological responses to emergencies and disasters.

1 Introduction

Evaluation is an essential part of health emergency preparedness exercises (HEPE) that allows to identify limitations in performance as well as practices that need to be reinforced and maintained (ECDC, 2014). Addressing limitations and practices that did not work well stimulates further improvements to prevent duplicating mistakes to enhance preparedness (Savoia, 2012).

However, the lack of methodologies to guide the selection of appropriate evaluation methods and tools is acknowledged in the emergency preparedness literature (Beerens and Tehler, 2016). Heterogeneity and inconsistency of exercise evaluation methods feature across different exercise types, with very fragmented and inconsistent approaches to collecting data from exercise participants (Skryabina et al., 2018, Williams et al., 2008).

Objective measures provided by external evaluators and observers are often used as the primary and only source of exercise evaluation data with focus given to evaluation of system level of preparedness and very rarely to individuals' learning from exercises. Collecting data from exercise participants has been strongly advocated (Evaluation of Exercises Handbook, 2011) and can be used to triangulate evidence and therefore contribute to a more rigorous exercise evaluation (Biddinger et al., 2010). However, based on the reported literature, it is not routine practice. Particular concerns about self-assessment relate to creating over-inflated confidence, as exercise participants tend to look more favourably on their own behaviours (ECDC, 2014). In contrast, good consensus between self-assessment measures and evaluation conducted by external observers was reported (Biddinger et al., 2010, Savoia et al., 2009b) and participants were found to be more critical of self-performance than objective measures of their performance produced by external evaluators (Freimuth et al., 2008). The lack of valid and reliable tools to assist with participants' evaluation has also been acknowledged (Skryabina et al., 2018).

Participants' contribution to exercise evaluation may include feedback on the exercise conduct, their perceptions of the emergency response system capabilities, and perceptions of their learning from the exercise (Skryabina et al., 2018). Routinely collecting participant feedback has been recommended for any exercise type, and particularly for exercises designed for accountability purposes; to exclude the likelihood of poorly designed exercises affecting organisational performance during the exercise (Skryabina et al., 2018). Factors that affect performance in an emergency preparedness exercise were discussed earlier (Skryabina et al., 2017), and collecting feedback from participants against these criteria could contribute to improved design and delivery of future exercises.

Attempts have been made to create evaluation tools for participants' self-assessment of organisational capabilities from an exercise, including questionnaires validated for both discussion-based (Savoia et al., 2009b) and operation-based exercises (Savoia et al., 2010). However published examples of exercises where participants are involved in providing their own assessments of organisational performance from an exercise are rare, and the use of external evaluators remains the gold standard (Savoia et al., 2010).

Routinely collecting evaluation data from participants can not only contribute to exercise evaluation, but also enhance the educational value of an exercise at an individual level. Self-evaluation pre- and post-exercise engages participants in an act of self-reflection which, can stimulate the development of meta-cognitive knowledge (Ford and Schmidt, 2000). Metacognitive knowledge heightens learners' awareness of their own ability (or lack of), providing the potential for learners to act on this, both pre-and post-exercise. Therefore, self-evaluation pre-exercise can prompt or motivate participants to prepare for, or engage with the exercise, and self-evaluation post-exercise may motivate participants to undertake necessary actions to improve preparedness (Anderson, 1998). In addition, pre-exercise assessment is thought to highlight the key concepts and objectives in advance of an exercise (Kotora et al., 2014, Fowkes et al., 2010) which serves to focus exercise participants' attention to relevant details (Johnstone et al., 1994).

Participant self-assessment, typically conducted in exercises perusing learning purposes, has been used to demonstrate that exercise participation changed participants' perceptions of their own and organisational emergency preparedness (Beaton et al., 2003, Stergachis et al., 2007); developed understanding of their team capabilities and capabilities of their response partners (Emery et al., 2009, High et al., 2010, Morris et al., 2012, Perry, 2004); improved their competency-based knowledge and skills (Henning et al., 2004, Savoia et al., 2009a); and reduced perceived level of risk associated with their roles in emergency response (Peterson and Perry, 1999). However, evaluation tools used in those studies were mainly in-house designed questionnaires with no further information on their psychometric properties. The lack of valid and reliable tools to assist with participants' evaluation has also been earlier acknowledged (Skryabina et al., 2018).

To address this gap, and to increase an awareness about the importance of routinely collecting data from exercise participants about their experience of HEPE, this paper reports on work done by the authors to develop and test the reliability and validity of a tool to measure participants feedback, perceptions of emergency preparedness, and satisfaction with a preparedness exercise. Kirkpatrick's model (Kirkpatrick, 1998) for the evaluation of training programs was used as a framework for the design of the satisfaction scale of this questionnaire. The model utilises four levels of potential outcomes that can be measured from an educational intervention (such as an emergency preparedness exercise) which include: Reactions (how satisfied are participants with the intervention); Learning (what knowledge or skills have been learned from an exercise); Behaviour (what changes in behaviour have occurred as a result of exercise participation) and Results (what impact does an exercise have on the outputs of the system).

The questionnaire design involved identification of key constructs and outcomes; the validation of items; a scale construct; and checks for scale reliability, discriminant validity, and sensitivity to change. Feedback from emergency preparedness experts on the questionnaires' applicability and utility as an exercise evaluation tool was collected.

2 Methods

2.1 Identification of key outcomes

Outcome variables were selected via literature reviews and discussions with emergency preparedness specialists and using the framework structure offered by Kirkpatrick's model (Kirkpatrick, 1998) to study participant's reactions, self-reported learning, and behavioural intentions post-exercise. The participant pre-exercise preparation subscale was added based on needs identified through the analysis of educational literature and literature related to the perceived effectiveness of emergency exercises. A central concern was to identify the range of variables which would support construction of each of these scales.

Most items identified for the scales on perceptions, post-exercise feedback, and personal satisfaction were adapted from the studies reviewed in a recent extensive literature review of emergency preparedness exercises (Skryabina et al., 2017). The items for the pre-exercise subscale were mainly constructed based on the analysis of factors affecting attitudes and behaviours, as well as cognitive engagement with a task.

2.1.1 Pre-exercise preparation scale

The pre-exercise scale was intended to help participants to reflect on their attitudes towards an exercise, assess their current level of preparedness and the value of having personal objectives for an exercise, and to check for completion of the pre-exercise requirements outlined by the exercise planners. These questions may stimulate participants' cognitive perception filters (Johnstone et al., 1994), which control attention and engagement with exercise tasks, thus enhancing the potential educational benefits from exercise participation. Responses to these questions can also provide exercise planners with valuable information on whether participants have sufficient time to prepare for the exercise, or whether they understand and share its value. This could help with planning and engaging exercise participants for future exercises.

2.1.2 Participants' perception scale

Outcome variables for the participants' perception scale were mainly selected based on the content analysis of the papers discussed in the earlier reported review of emergency exercise effectiveness (Skryabina et al., 2017) and consultation with subject matter experts, comprising Public Health England (PHE) Emergency Preparedness Resilience and Response (EPRR) managers and EPRR training specialists, and also a diverse group of stakeholders including hospital emergency preparedness managers, clinicians, emergency preparedness exercises managers, behavioural scientists, a military emergency response training coordinator and a PHE Emergency Response Department (ERD) communications officer. Contribution to the questionnaire content was also sought from two public representatives of the project advisory group. Applying the perceptions scale pre- and post-exercise can clarify whether any learning was achieved from exercise participation.

2.1.3 Post-exercise feedback scale

The post-exercise feedback scale consists of items identified from the content analysis of the reviewed literature, as do the other scales. The content analysis identified that the quality of exercise facilitators, presence of the key participants involved in the response, functional fidelity provided by the exercise scenario, and sufficient time to complete the task were the strongest factors affecting the exercise outcomes. The importance of setting up an objective evaluation criteria and providing suitable facilities to accommodate the exercise needs were also mentioned in some studies. All these factors were included as items for the exercise feedback scale.

2.1.4 Personal satisfaction scale

Kirkpatrick's model (Kirkpatrick, 1998) for the evaluation of training programs was used as a framework for the design of the personal satisfaction scale, which measures participants' overall satisfaction with the exercise; any improvements in perceptions of their personal and organisational level of emergency preparedness; behavioural intentions; and the capacity to engage in behaviour change post-exercise. These last two items are important for any follow-up studies to verify whether behavioural intentions led to change in behaviour and practices as they relate to real incident response.

2.2 Scale construct

To make it easy for respondents, the same type of scale, namely a six-point Likert scale, was used for all subscales of the project questionnaire. The scale was used, with the following options: Strongly agree (6), Agree (5), More Agree than Disagree (4), More Disagree than Agree (3), Disagree (2), Strongly Disagree (1). Taking into account that not all aspects of the emergency response and emergency plans are relevant to all range of emergency response roles, an additional option of N/A (Not Applicable) was offered. After consideration of the appropriateness of using a scale without a clear midpoint (Leung, 2011), the final decision of offering a 6-point scale was based on the grounds that the considered topic has a relevance to any emergency responder and that they should have an opinion and be able to articulate their perceptions and views by using a scale without a neutral point. The theory of attitude suggests that personal relevance of a topic generates more polarised views, while a personally unrelated object produces neutral evaluations (Petty and Cacioppo, 1981). It would be appropriate to expect that emergency preparedness-related questions will be relevant for healthcare emergency responders. In fact, as a part of the outcome of emergency exercises, it is desirable for participants to formulate their positions and perceptions in relation to the topics considered.

2.3 Validation of items

The items for all four scales were validated first in two focus groups: the first group consisted of four behavioural scientists and the second group included five experienced emergency preparedness exercise managers. Each focus group was asked to complete the questionnaire first and note any identified issues; afterwards each item was discussed to check for its validity, language clarity and appropriateness of the 6-point response scale.

Most of the comments collected at this stage were addressed in the next version of the questionnaire. Further feedback was requested from the members of the project advisory group, including emergency preparedness healthcare and military specialists, clinicians and members of the public. They were requested to comment on the questionnaire content accuracy, appropriateness of items, readability, and appropriateness of the scale selection. In total, 19 experts took part in the questionnaire item validation. All recommendations were reviewed by the research team, and appropriate changes were introduced in the final version of the questionnaire, which was further piloted twice with a range of healthcare and emergency preparedness professionals. Two pilot studies took place two weeks apart, and the data obtained was used for the questionnaire reliability check.

Ethics permission was not required for these pilot studies as participants were healthcare professionals involved in the study by virtue of their roles. This position was confirmed by the PHE Research Ethics and Governance Group (REGG).

Participants of four large regional discussion-based tabletop exercises (TTX) and four large English regional Major Trauma Network (MTN) operational exercises were then invited to complete the questionnaire pre- and post-exercise. Ethics permission to collect the data from discussion-based exercise participants for research purposes was granted by the E&M Research Ethics Panel, King's College London, UK (LRS-16/17-4611 from 3.05.2017). Permission to collect data from participants of operation-based exercises was granted by the University of Liverpool Institute of Psychology, Health and Society Ethics Committee (IPHS-1213-SG-052 – 5th Amendment from 14.12.17).

In addition, opinions of emergency preparedness experts, including exercise managers, were solicited and collected on the tool's usability and utility.

2.4 Statistical analysis

Test-retest reliability of the items was assessed using Interclass Correlation Coefficient (ICC). The scales' internal reliability was checked by using Cronbach's Alpha and inter-item correlations. The scale discriminant validity was tested by comparing pre-exercise perceptions of less experienced staff (fewer than four years in practice) with perceptions of more experienced staff (more than nine years in practice). Scale sensitivity was checked by comparing pre-exercise to post-exercise perceptions of less experienced staff (fewer than four years in practice). IBM SPSS Statistics for Windows, Version 24.0 (Armonk, NY: IBM Corp) was used for the statistical analysis.

3 Results

The first stage of the study included collecting data for test-retest analysis from EPRR experts and participants of a regional discussion-based TTX ($N = 27$). The second stage involved collecting data from participants of four regional discussion-based exercises ($N = 97$) and four operational exercises ($N = 238$) to study the questionnaire's internal consistency, discriminant validity and sensitivity to change. Additional data was collected from 13 EPRR experts to study the questionnaire's usability and utility.

3.1 Sample

Demographic characteristics of 27 participants who took part in the questionnaire test-retest study are given in Table 1. Most EPRR experts (16, 59.3%) were from Public Health England (PHE) and their daily jobs directly related to emergency preparedness and response, including emergency planning managers (8, 29.6%), medical health protection staff (5, 18.5%), emergency preparedness exercises managers (2, 7.4%), regional leads of emergency preparedness resilience and response (2, 7.4%) and a business support manager (1, 3.7%). Most experts were between 50–59 (11, 40.7%) or between 30–39 (8, 29.6%) years of age, with an average of 3.7 years of practice in their current roles. The gender distribution of the group was balanced: 14 reported gender identification as female (51.8%) and 13 as male (48.2%).

Table 1 Demographic characteristics of participants in the test-retest reliability study ($N = 27$)

	N = 27, Frequency %
Age	
18–29	2 (7.4%)
30–39	8 (29.6%)
40–49	4 (14.8%)
50–59	11 (40.7%)
60+	2 (7.4%)
Gender	
Female	14 (51.8%)
Male	13 (48.2%)
Organisation	
Public Health England (PHE) EPRR	16 (59.3%)
Military	2 (7.4%)
Hospital (EPRR)	5 (18.5%)
Ambulance	1 (3.7%)
NHS England	1 (3.7%)
Other	2 (7.4%)
Day Role	
Business Support Manager	1 (3.7%)
Emergency Planning Manager	8 (29.6%)
Emergency Exercise Manager	2 (7.4%)
Medical (Health Protection)	5 (18.5%)
EPRR Regional Leads	2 (7.4%)
TTX participant	9 (33.3%)
Day Role (Category)	
Clinical	1 (3.7%)
EPRR	16 (59.2%)
Managerial	7 (25.9%)
Other	4 (14.8%)
Years of practice, Median [range]	2.0 [0–18]

The data to check the questionnaire scales' internal reliability, discriminant validity, and scale sensitivity was collected from healthcare staff who took part in both, discussion-based tabletop exercises (TTX) and operation-based major trauma network (MTN) emergency preparedness exercises, and who completed the questionnaire pre-exercise [TTX ($N = 97$); MTN ($N = 238$)] and post-exercise [TTX ($N = 93$); MTN ($N = 95$)]. Participants' demographic characteristics are given in Table 2. Almost half of TTX participants were NHS Trusts (Hospital) staff (46, 47.4%), with operational staff (28; 28.9%) and support staff (26; 26.8%) dominating. However, there was also a good representation of tactical (24, 24.7%) and strategic (16, 16.5%) emergency response roles, with a balance of participants identifying as female (51, 52.6%) and male (40, 41.2%) respectively. TTX participants from other organisations included those from Clinical Commissioning Groups (CCGs), Social Care, Local Authorities, NHS Blood and Transplant, the Department of Health and Social Care, and Fire and Rescue Services. The prevalent age group for TTX was 40–49 (38, 39.2%), with two other groups of 30–39 (23, 23.7%) and 50–59 (25, 25.8%) being almost equally represented. Participants had a diverse range of work experience (median 3 years, range [0–25 years]).

Table 2 Demographic characteristics of study participants from discussion-based tabletop exercises (TTX), $N = 97$ and operation-based exercises (MTN), $N = 238$

	TTX (N = 97)	*MTN (N = 238)*
Item	*Frequency (%)*	*Frequency (%)*
Age		
18–29	2 (2.1%)	13 (5.5%)
30–39	23 (23.7%)	45 (18.9%)
40–49	38 (39.2%)	79 (33.2%)
50–59	25 (25.8%)	82 (34.5%)
60+	4 (4.1%)	15 (6.3%)
Gender		
Female	51 (52.6%)	143 (60.1%)
Male	40 (41.2%)	95 (39.9%)
Organisation		
Ambulance	6 (6.2%)	31 (13%)
NHS Trust (Hospital)	46 (47.4%)	172 (72.3%)
NHS England	16 (16.5%)	11 (4.6%)
Public Health England (PHE)	2 (2.1%)	–
Other (support)	27 (27.8%)	24 (10.1%)
Day Role (Category)		
Clinical	20 (20.6%)	89 (37%)
EPRR	17 (17.5%)	23 (9.7%)
Managerial	40 (41.2%)	73 (30.7%)
Other	18 (18.6%)	53 (22.6%)

Table 2 Demographic characteristics of study participants from discussion-based tabletop exercises (TTX), $N = 97$ and operation-based exercises (MTN), $N = 238$ (continued)

Item	*TTX (N = 97)*	*MTN (N = 238)*
	Frequency (%)	*Frequency (%)*
Emergency Role (Category)		
Operational (Bronze)	28 (28.9%)	153 (64.3%)
Tactical (Silver)	24 (24.7%)	61 (25.6%)
Strategic (Gold)	16 (16.5%)	24 (10.1)
Other	26 (26.8%)	–
Years of practice, Median [range]	3 [0–25]	4 [0–40]

Participants of operation-based MTN exercises were mainly represented by NHS Trusts (Hospital) staff (172, 72.3%), with operational emergency roles dominating (153, 64,3%). Almost a quarter of participants had a tactical role (61, 25.6%), with the majority identifying as female (143, 60.1%). Ambulance personnel comprised 31 (13%) participants, with a relatively small representation from other health organisations. Most participants were between 30–59 years, with a diverse range of work experience (median 4 years, range [0–40 years]); (Table 2).

The discussion-based exercises that participants attended were regional tabletop exercises aimed at exploring the recovery stage from a large-scale mass casualty major incident in England. Participants from a variety of healthcare organisations and local authorities which would likely be involved in a major incident recovery stage were invited to take part in each exercise. The operation-based exercises involved into this study aimed at practicing major trauma network (MTN) response to a marauding terrorist firearms attack (MTFA) major incident involving a large number of casualties. The exercise used the simulation casualty data from the Emergo Train System® (ETS) (ETS, 2016). Characteristics of each exercise are given in the Table 3.

3.2 Reliability of questions

The questionnaire test-retest reliability was checked in two rounds of online surveys with subject matter experts and participants of a discussion-based exercise. A total of 27 responses were received from experts from the first survey, but only 18 of these responded on the second survey two weeks later. A few modifications were suggested and introduced to the questionnaire after the first survey, including re-wording existing questions to improve the clarity and introducing a few new questions. The majority of responders expressed their overall satisfaction with the content and the scale itself after the second survey. The second round of the data collection for the reliability study was undertaken with participants of a regional tabletop exercise. Only nine out of the 73 (12.3%) exercise participants, who were individually invited to take part in the reliability study, returned responses to two surveys two-weeks apart post-exercise. Reliability analysis is based on the data collected from both 18 EPRR experts and nine exercise participants (total $N = 27$).

Table 3 Exercises characteristics: regional discussion-based tabletop exercises (TTX) and regional major trauma network (MTN) operation-based exercises; 2017–2018

Exercise	*NHS England region*	*Number of exercise participants*	*Recruited study participants*	*Scenario; N of casualties and fatalities*
Salus (2017)	South	72	24 (33%)	MTFA; 1600 casualties
Alamein (2017)	Midlands and East	64	20 (31%)	Train derailment; 540 casualties and 25 fatalities
Seacole II (2017)	London	41	11 (27%)	A vehicle-borne attack; 458 casualties and 86 deaths
Stonehart (2017)	North	162	54 (33%)	Incident at an ice rink, with over 300 casualties, including burns patients and 34 fatalities and Road Traffic Collision (RTC) with over 50 casualties (elderly)
Tartar (2018)	Midlans and East Trauma Network	219	36 (16%)	A tram crash incident and a MTFA in a shopping centre, up to 292 casualties
Golden Eagle (2018)	Cheshire and Mersey Trauma Network	409	80 (20%)	A tram crash incident and a MTFA in a shopping centre, up to 232 casualties
Kestrel (2018)	Sussex Trauma Network	124	57 (46%)	MTFA with subsequent bombing and edged weapon/knife attack; presenting 252 trauma patients
Blue Peter (2018)	Midlands and East Trauma Network	238	124 (52%)	Two MTFAs with subsequent bombings and edged weapon/knife attacks; presenting up to 341 trauma patients

Interclass correlation coefficient (ICC) for some items from the Pre-exercise preparation scale was rather low (0.413–0.619); the same was true for a few Post–exercise feedback scale subscales including aims and objectives (0.201–0.404), exercise scenario (–0.375 (item 54) and 0.033 (item 56)) and exercise format (–0.235–0.118). Very good agreement on aims and objectives and exercise scenario subscale items were obtained from the nine discussion-based exercise participants only [0.632–1.000]. However, this agreement did not extend to the exercise format subscale. The test-retest agreement was good for most subscales from the perceptions scale, and particularly for the perceptions of emergency plans (0.816–0.966), except perceived confidence in organisation's hazard specific plans (ICC = 0.373; item 32); perceptions of organisational preparedness (0.799–0.858); and for perceived level of stress (0.719–0.726). Similarly, agreement was good for some subscales of the Post-exercise scale, including key players (0.779–0.847); exercise time (0.730–0.919); evaluation (0.785–0.927); venue (0.888); exercise facilitators (0.688 – 0.954) [with the exception of perceptions of facilitator help in keeping discussions to time (ICC = 0.500; item 62)]; and the personal satisfaction with the exercise subscale

(0.710–0.931), [with the exception of the item assessing the capability of addressing training needs highlighted by the exercise (ICC = 0.544, item 85)]. Results of the test-retest analysis (ICC) are shown in Table 4.

Table 4 Reliability analysis data presented separately for discussion-based (TTX) and operation-based (MTN) exercises

N	*Item*	*ICC, N = 27*	*Exercise type*	*Cronbach's Alpha*	*Cronbach's Alpha when deleted, N = 93*	*Mean, max 6*
	PRE-EXERCISE TTX (N = 93)		TTX	*0.883*		*4.74*
	MTN (N = 238)		MTN	*0.879*		*4.91*
1	I am motivated to take part in the exercise	0756	TTX		0.877	5.33
			MTN		0.870	5.35
2	I understand the aim of the exercise	0.530	TTX		0.867	5.09
			MTN		0.863	5.10
3	I understand the exercise objectives	0.574	TTX		0.866	4.84
			MTN		0.864	4.92
4	The exercise objectives are relevant to my emergency role	0.619	TTX		0.870	4.85
			MTN		0.871	5.05
5	The exercise objectives are relevant to my every day role	0.673	TTX		0.870	4.34
			MTN		0.870	4.76
6	The exercise objectives are relevant to my organisation	0.634	TTX		0.875	5.24
			MTN		0.873	5.37
7	I think this exercise is timely	0.333	TTX		0.880	5.35
			MTN		0.871	5.07
8	I expect that the exercise will be useful for me, personally	0.753	TTX		0.868	5.11
			MTN		0.869	5.32
8a	I understand my role in the exercise	0.600	TTX		–	–
			MTN		0.870	4.73
9	I have had sufficient time to prepare for this exercise	0.674	TTX		0.877	3.94
			MTN		0.872	4.42
10	I have identified my personal objectives for this exercise	0.413	TTX		0.870	3.89
			MTN		0.863	4.25
11	I completed the pre-exercise preparations required by the exercise team**	0.152	TTX		0.889	3.74
			MTN		0.885	4.25
12	I understand the value of this exercise	0.597	TTX		0.870	5.10
			MTN		0.866	5.26

Table 4 Reliability analysis data presented separately for discussion-based (TTX) and operation-based (MTN) exercises (continued)

N	*Item*	*ICC, N = 27*	*Exercise type*	*Cronbach's Alpha*	*Cronbach's Alpha when deleted, N = 93*	*Mean, max 6*
	Perceptions (pre-exercise)TTX (N = 93); MTN (N = 238)					
	Perception of training and preparedness	*ICC*	*TTX*	*0.885*		*4.24*
			MTN	*0.846*		*4.46*
13	The training I have received prepared me well to respond to a major incident	0.676	TTX		0.831	4.08
			MTN		0.786	4.38
14	I understand the requirements of my role in a major incident	0.859	TTX		0.859	4.53
			MTN		0.777	4.87
15	I am confident in my abilities to operate effectively in a major incident	0.619	TTX		0.859	4.52
			MTN		0.803	4.71
16	I have practiced my emergency role in emergency preparedness exercises	0.600	TTX		0.859	3.84
			MTN		0.867	3.87
	Perceptions of team work	*ICC*	*TTX*	*0.893*		*4.65*
			MTN	*0.856*		*4.79*
17	I have confidence in my team's ability to respond in a major incident	0.509	TTX		0.828	4.65
			MTN		0.770	4.75
18	I believe my team members are competent in their response roles	0.593	TTX		0.861	4.54
			MTN		0.761	4.70
19	I believe my team knows where to get support in a major incident	0.749	TTX		0.840	4.53
			MTN		0.795	4.71
20	I feel I am a valued member of my team **	0.699	TTX		0.913	4.89
			MTN		0.910	4.98
	Perceptions of resources	*ICC*	*TTX*	*0.695*		*3.76*
			MTN	*0.780*		*4.09*
21	Our department has the necessary equipment to deal with a large number of casualties	0.791	TTX		0.571	2.97
			MTN		0.678	3.82
22	I have the right equipment to perform my role in a major incident	0.537	TTX		0.689	4.05
			MTN		0.763	4.46
23	Our department is able to maintain the supply of necessary resources in response to a major incident	0.369	TTX		0.497	3.65
			MTN		0.670	4.00
24	Our organisation understands how to request mutual-aid support**	0.107	TTX		0.713	4.36
			MTN		0.783	4.16

Table 4 Reliability analysis data presented separately for discussion-based (TTX) and operation-based (MTN) exercises (continued)

N	*Item*	*ICC, N = 27*	*Exercise type*	*Cronbach's Alpha*	*Cronbach's Alpha when deleted, N = 93*	*Mean, max 6*
	Perceptions of multi-agency response	*ICC*	*TTX*		*0.819*	*4.47*
			MTN		*0.865*	*4.52*
25	I understand other agencies' roles during a major incident	0.763	TTX		0.794	4.46
			MTN		0.858	4.56
26	I think other agencies have a good understanding of my organisation's role	0.716	TTX		0.822	4.23
			MTN		0.839	4.42
27	I am confident that other responding organisations can work effectively in a major incident response	0.901	TTX		0.776	4.55
			MTN		0.818	4.50
28	I am confident that my organisation can effectively work together with other responding organisations in response to a major incident	0.597	TTX		0.765	4.77
			MTN		0.845	4.73
29	I am confident that responding organisations have developed shared understanding about the response process and strategies	0.132	TTX		0.764	4.35
			MTN		0.818	4.37
	Perceptions of emergency plans	*ICC*	*TTX*	*0.870*		*4.41*
			MTN	*0.880*		*4.56*
30	I have confidence in our departmental (local) incident plan	0.969	TTX		0.839	4.31
			MTN		0.854	4.62
31	I have confidence in our organisation's incident plan	0.898	TTX		0.846	4.48
			MTN		0.852	4.69
32	I have confidence in our organisation's hazard specific plans	0.373	TTX		0.845	4.03
			MTN		0.862	4.28
33	I have confidence in the national major incident plan	0.910	TTX		0.869	4.02
			MTN		0.866	4.47
34	I am confident that my department/team plan complements the organisations incident plan	0.519	TTX		0.849	4.07
			MTN		0.876	4.58
35	I am confident that my local plan complements the national plan	0.532	TTX		0.852	4.21
			MTN		0.859	4.48
36	I believe that my organisation's emergency plans are regularly reviewed and updated	0.816	TTX		0.857	4.67
			MTN		0.866	4.79

Table 4 Reliability analysis data presented separately for discussion-based (TTX) and operation-based (MTN) exercises (continued)

N	*Item*	*ICC, N = 27*	*Exercise type*	*Cronbach's Alpha*	*Cronbach's Alpha when deleted, N = 93*	*Mean, max 6*
	Perceptions of organisational preparedness	*ICC*	*TTX*	*0.906*		*4.35*
			MTN	*0.857*		*4.48*
37	I am confident that my organisation can respond effectively in a major incident.	0.858	TTX		0.869	4.65
			MTN		0.821	4.69
38	My department is well prepared to respond in a major incident	0.915	TTX		0.902	4.27
			MTN		0.833	4.52
39	My organisation is prepared for a multi-agency response	0.799	TTX		0.877	4.41
			MTN		0.806	4.55
40	Emergency preparedness is a high priority in my organisation	0.857	TTX		0.887	4.34
			MTN		0.816	4.49
41	My organisation regularly participates in multi-agency training and exercising	0.806	TTX		0.891	4.09
			MTN		0.862	4.15
	Perceived level of stress	*ICC*	*TTX*	*0.452*		*3.92*
			MTN	*0.492*		*3.56*
42	I feel anxious when thinking about taking part in a major incident response	0.726	TTX			3.38
			MTN			3.56
43	If a major incident occurred today I would feel confident to take part in the response	0.719	TTX			4.46
			MTN			4.61
	Perceptions of competency-based knowledge and skills	*ICC*	*TTX*	*0.906*		*4.64*
			MTN	*0.905*		*4.73*
44	I can describe my functional role(s) and responsibilities in a major incident	0.810	TTX		0.900	4.51
			MTN		0.898	4.79
45	I can describe my organisation's role in a major incident	0.767	TTX		0.890	4.62
			MTN		0.880	4.79
46	I can describe my organisation's coordination (chains of command) in a major incident	0.826	TTX		0.882	4.51
			MTN		0.882	4.69
47	I can locate my organisation's incident response plan	0.440	TTX		0.892	5.00
			MTN		0.898	5.02

Table 4 Reliability analysis data presented separately for discussion-based (TTX) and operation-based (MTN) exercises (continued)

N	*Item*	*ICC, N = 27*	*Exercise type*	*Cronbach's Alpha*	*Cronbach's Alpha when deleted, N = 93*	*Mean, max 6*
48	I can describe the role of my organisation in a major incident response in relation to other organisations	0.897	TTX		0.884	4.55
			MTN		0.888	4.53
49	I am familiar with my organisation's preparedness plans and emergency arrangements	0.844	TTX		0.886	4.64
			MTN		0.880	4.58
	Post exercise feedback					
	Aim and objectives	*ICC*	*TTX*	*0.816*		*4.91*
			MTN	*0.729*		*5.06*
50	The exercise aim was achieved	0.404	TTX		0.817	4.86
			MTN		0.647	4.92
51	The exercise objectives were relevant to my emergency role	0.201	TTX		0.636	4.77
			MTN		0.667	4.87
52	The exercise objectives were relevant to my organisation	0.380	TTX		0.744	5.11
			MTN		0.692	5.38
	Exercise scenario	*ICC*	*TTX*	*0.878*		*5.02*
			MTN	*0.672*		*5.21*
53	The challenges presented by exercise were relevant to my organisation	0.018	TTX		0.854	5.09
			MTN		0.352	5.30
54	The scenario triggered actions related to the exercise objectives	–0.375	TTX		0.822	5.08
			MTN		–	–
55	The exercise scenario triggered actions that were relevant to my response roles and responsibilities	0.364	TTX		0.841	4.75
			MTN		0.829	4.99
56	The scenario presented challenges that facilitated learning	0.033	TTX		0.855	5.14
			MTN		0.501	5.35

Table 4 Reliability analysis data presented separately for discussion-based (TTX) and operation-based (MTN) exercises (continued)

N	*Item*	*ICC, N = 27*	*Exercise type*	*Cronbach's Alpha*	*Cronbach's Alpha when deleted, N = 93*	*Mean, max 6*
	Exercise format	*ICC*	*TTX*		*0.799*	*4.71*
			MTN		*0.803*	*4.44*
57	The exercise ground rules were clearly explained to me	–0.235	TTX		0.782	5.32
			MTN		0.814	4.99
58	The evaluation process was clear to me	0.118	TTX		0.742	5.01
			MTN		0.743	4.73
59	The facilitator's roles were explained to me	0.184	TTX		0.766	5.24
			MTN		0.761	5.01
60	I know when the exercise report will be available for me	–	TTX		0.804	3.98
			MTN		0.765	3.38
61	I understand how the lessons identified in this exercise will be actioned	0.946	TTX		0.712	4.01
			MTN		0.729	4.10
	Exercise facilitators ##	*ICC*	*TTX*	*0.940*		*4.63*
62	The facilitator was helpful in keeping discussion to time	0.500	TTX		0.927	4.86
			MTN			
63	The facilitator was helpful in keeping discussion on topic	0.721	TTX		0.927	4.79
			MTN			
64	All participants had an opportunity to participate in group discussions	0.760	TTX		0.948	5.24
			MTN			
65	The facilitator was good in identifying key points in discussion	0.688	TTX		0.925	4.82
			MTN			
66	The facilitator enabled effective group problem solving activities to take place	0.758	TTX		0.928	4.54
			MTN			
67	The facilitator was good in encouraging participants to refer to their plans to support discussions	0.943	TTX		0.929	4.12
			MTN			
68	The facilitator was helpful in practising a process of coming to a collectively-agreed response	0.864	TTX		0.933	4.29
			MTN			
69	The facilitator encouraged participants to actively seek out information from other participants within the room	0.954	TTX		0.937	4.39
			MTN			

Table 4 Reliability analysis data presented separately for discussion-based (TTX) and operation-based (MTN) exercises (continued)

N	*Item*	*ICC, N = 27*	*Exercise type*	*Cronbach's Alpha*	*Cronbach's Alpha when deleted, N = 93*	*Mean, max 6*
	Key players	*ICC*	*TTX*	*0.847*		*4.54*
			MTN	*0.735*		*4.46*
70	Key organisations that would be involved in a major response were represented at the exercise	0.824	TTX		0.793	4.72
			MTN		0.815	4.34
71	Functional leaders who would direct the organisation's response participated in the exercise	0.779	TTX		0.801	4.54
			MTN		0.576	4.67
72	Adequate representation of roles was present or accessible	0.847	TTX		0.764	4.37
			MTN		0.553	4.37
	Exercise time	*ICC*	*TTX*	*0.755*		*4.74*
73	There was sufficient time devoted during the exercise to consider the issues raised	0.730	TTX		0.748	4.78
74	Exercise injects were delivered at the right time	0.919	TTX		0.626	4.74
75	The time pressure created at the exercises was appropriate	0.836	TTX		0.614	4.71
	Evaluation	*ICC*	*TTX*	*0.750*		*4.47*
			MTN	*0.667*		*4.68*
76	The evaluation process was clear to me	0.895	TTX		0.655	4.51
			MTN		–	–
77	The evaluation methodology was objective	0.927	TTX		0.722	4.20
			MTN		0.802	4.05
78	Post-exercise hot-debrief identified important lessons	0.823	TTX		0.720	3.99
			MTN		0.362	4.94
79	I was able to share my feedback on the performance in the exercise	0.785	TTX		0.677	4.36
			MTN		0.590	5.06
	Venue	*ICC*				
80	The site/facilities met the needs of the exercise	0.888	TTX			5.28

Table 4 Reliability analysis data presented separately for discussion-based (TTX) and operation-based (MTN) exercises (continued)

N	*Item*	*ICC, N = 27*	*Exercise type*	*Cronbach's Alpha*	*Cronbach's Alpha when deleted, N = 93*	*Mean, max 6*
	Personal satisfaction		*TTX*	*0.864*		*4.68*
			MTN	*0.840*		*4.85*
81	I learned something new from the exercise	0.878	TTX		0.848	5.13
			MTN		0.826	5.33
82	I feel more confident to respond to a major incident after this exercise	0.931	TTX		0.846	4.83
			MTN		0.825	5.01
83	The exercise has identified gaps in my emergency preparedness knowledge/training	0.862	TTX		0.844	4.46
			MTN		0.833	4.57
84	I am motivated to improve my emergency preparedness knowledge	0.887	TTX		0.852	5.13
			MTN		0.819	5.11
85	I am capable of addressing my training needs highlighted by this exercise	0.544	TTX		0.847	4.69
			MTN		0.814	4.87
86	I will translate the learning from this exercise to my day job	0.833	TTX		0.849	4.74
			MTN		0.838	4.36
87	I am confident that the lessons identified will be addressed and embedded	0.842	TTX		0.854	4.53
			MTN		0.827	4.53
88	I expect to be advised how the lessons identified in the exercise will be addressed	0.854	TTX		0.861	4.44
			MTN		0.831	4.72
89	I understand who will lead on ensuring the lessons identified will be actioned	0.837	TTX		0.859	3.97
			MTN		0.841	4.55
90	I would recommend this type of exercise for my colleagues	0.747	TTX		0.851	5.12
			MTN		0.828	5.34
91	I will share learning from this exercise with my colleagues	0.710	TTX		0.852	5.08
			MTN		0.823	5.00
92	Organisational limitations identified in this exercise are beyond my control	0.875	TTX		0.882	4.03
			MTN		–	–

**items can be removed from the final questionnaire to improve the scale internal reliability.

Facilitator subscale was only offered for TTX exercise participants.

The scale internal consistency was checked using Cronbach's alpha correlation coefficient. Checks were undertaken separately for discussion-based TTX and operation-based MTN exercises (Table 4). Except for the subscale on the perceived level of stress

[Cronbach's alpha = 0.452 (TTX) and 0.492 (MTN)], all other scales had high internal consistency for both types of exercises (Cronbach's alpha = 0.672–0.940). Three individual items, which improve the Cronbach's alpha for both TTX and MTN exercises when deleted, were identified (11, 20, 24) and are removed from the final questionnaire. Following this analysis it was also decided to eliminate the perceived level of stress subscale, and to consider its two questions as individual items ("*I feel anxious when thinking about taking part in a major incident response*" and "*If a major incident occurred today I would feel confident to take part in the response*").

3.3 Discriminant validity

Discriminant validity was checked by using an Independent Samples t-test or its non-parametric alternative (the Mann-Whitney test) to analyse differences in perceptions of emergency preparedness for less-experienced staff (fewer than four years of practice) and more experienced staff (more than nine years of practice) pre-exercise, for both exercises types, discussion-based TTX (Table 5(a)) and operation-based MTN exercises (Table 5(b)). For TTX participants, perceptions of competency-based knowledge and skills, teamwork, appropriateness of resources, perceptions of multi-agency response, organisational preparedness and emergency plans did not differ between more- and less-experienced staff, but there were significant differences in the perceptions of training and preparedness. Less-experienced staff also felt significantly more anxious when thinking about taking part in a major incident, as well as felt less confident to take part in a response if a major incident occurred today, compared with their more-experienced colleagues.

Table 5(a) Discriminant validity analysis (Perception scale, pre-exercise), discussion-based tabletop exercises (TTX)

Subscale	*Mean, max*	*Group 1 (N = 52), Mean*	*Group 2 (N = 22) Mean*	*Sig. (2-tailed)*
Perception of training and preparedness	24	15.75	19.36	0.001
Perceptions of team work	24	18.42	19.81	0.07
Perceptions of appropriateness of resources	24	14.44	16.22	0.161
Perceptions of multi-agency response	30	21.98	23.31	0.141
Perceptions of emergency plans	42	28.57	32.05	0.083
Perceptions of organisational preparedness	30	21.19	23.32	0.094
Perceptions of competency-based knowledge and skills	36	26.86	29.18	0.147
I feel anxious when thinking about taking part in a major incident response#	6	4.27	3.53	0.048
If a major incident occurred today I would feel confident to take part in the response#	6	4.22	5.00	0.004

- Group 1 – less than 4 years of experience; Group 2 – more than 9 years of experience.
- The difference is significant at equal or less than 0.05.
- #these are the single questions, not a subscale.

Table 5(b) Discriminant validity analysis (perception scale, pre-exercise), operation-based major trauma network (MTN) exercises

Subscale	*Mean, max*	*Group 1 (N = 43), Mean*	*Group 2 (N = 14) Mean*	*Sig. (2-tailed)*
Perception of training and preparedness	24	17.23	16.57	0.609
Perceptions of team work	24	18.84	17.57	0.271
Perceptions of appropriateness of resources	24	16.31	15.46	0.540
Perceptions of multi-agency response	30	22.74	21.38	0.311
Perceptions of emergency plans	42	32.10	30.36	0.251
Perceptions of organisational preparedness	30	22.14	21.86	0.835
Perceptions of competency-based knowledge and skills	36	27.33	26.86	0.807
I feel anxious when thinking about taking part in a major incident response#	6	3.98	3.93	0.912
If a major incident occurred today I would feel confident to take part in the response#	6	4.47	4.21	0.410

- Group 1 – less than 4 years of experience; Group 2 – more than 9 years of experience.
- The difference is significant at equal or less than 0.05.
- #these are the single questions, not a subscale.

However, no differences in any of the considered pre-exercise perceptions were found between less and more experienced participants in operation-based MTN exercises.

3.4 Sensitivity to change

The sensitivity analysis was only conducted for the perceptions scale, as it was the only scale offered both pre- and post- exercise. The scale sensitivity was checked by comparing the less-experienced participants' post-exercise perceptions of their personal, social (team and multi-agencies) and organisational levels of emergency preparedness to their pre-exercise perceptions. Significant differences ($p < 0.05$) for seven out of eight perception subscales indicated a significant positive change in less-experienced participants in both types of exercises, discussion-based TTX and operations-based MTN (Tables 6(a) and (b)), as would be expected from a well-designed and well-delivered exercise. No difference was observed for the 'perceptions of resources' subscale for TTX, and for the 'perceptions of multi-agency response' subscale with MTN exercise participants.

For comparison, a similar analysis was conducted with a group of less-experienced healthcare providers who did not attend any exercise over the past 6 months (Control group). For this group, the only significant difference in responses, collected two weeks apart, was in the perceptions of emergency plans; no other differences were observed. Participation in the research project might have prompted some of these control group participants to refresh their knowledge of emergency plans; this could explain the observed difference without an exercise attendance as well as to prove a positive effect of taking part in the survey.

Table 6(a) Perceptions scale sensitivity analysis for less experienced staff (less than 4 years of experience); Discussion-based TTX exercises

	Exercise group (TTX), N = 47			*Control group, N = 19*		
Subscale	*Pre-ex, Mean*	*Post-ex, Mean*	*Sig. (2-tailed)*	*Pre-ex, Mean*	*Post-ex, Mean*	*Sig. (2-tailed), P*
Perception of training and preparedness	16.30	19.13	0.00	15.16	16.16	0.20
Perceptions of team work	18.91	20.28	0.00	18.47	19.68	0.31
Perceptions of appropriateness of resources	14.47	15.75	0.19	14.00	15.47	0.13
Perceptions of multi-agency response	22.25	23.70	0.00	20.21	21.05	0.14
Perceptions of emergency plans	29.55	31.70	0.01	28.10	30.47	0.04
Perceptions of organisational preparedness	22.47	24.68	0.00	20.74	21.95	0.08
Perceptions of competency-based knowledge and skills	27.66	30.32	0.00	26.58	26.53	0.96
I feel anxious when thinking about taking part in a major incident response*	3.68	3.28	0.03	3.00	3.32	0.21
If a major incident occurred today I would feel confident to take part in the response*	4.34	4.94	0.00	4.00	4.00	1.00

Paired t-test; significance at 95% confidence, $p < 0.05$.

Table 6(b) Perceptions scale sensitivity analysis for less experienced staff (less than 4 years of experience, $N = 43$); Operation-based (MTN)

Subscale	*Mean max*	*Pre-ex, Mean*	*Post- ex, Mean*	*Sig. (2-tailed)*
Perception of training and preparedness	24	17.23	19.33	0.000
Perceptions of team work	24	18.84	19.91	0.001
Perceptions of appropriateness of resources	24	16.59	17.66	0.000
Perceptions of multi-agency response	30	22.74	22.77	0.980
Perceptions of emergency plans	42	32.10	33.07	0.000
Perceptions of organisational preparedness	30	22.14	23.07	0.000
Perceptions of competency-based knowledge and skills	36	27.33	30.65	0.000
I feel anxious when thinking about taking part in a major incident response*	6	3.98	3.56	0.000
If a major incident occurred today I would feel confident to take part in the response*	6	4.47	4.91	0.000

Paired t-test; significance at 95% confidence, $p < 0.05$.

3.5 Usability and utility

Feedback on the scales' usability and utility was received from 13 Emergency Preparedness Resilience and Response (EPRR) experts from NHS England and Public Health England. All have previously been involved in setting up bespoke evaluation methods for emergency exercises and have, on average, six years of experience in their roles (range 1–17 years).

Table 7 Utility analysis, percent of positive responses/Mean (5 point-scale)

Usability assessment items	*Pre-exercise scale, N = 13*	*Perceptions scale, N = 13*	*Exercise feedback, N = 9*	*Personal satisfaction with the exercise scale, N = 9*
The scale is of appropriate length	84.6%/4.00	23.1%/2.75	88.9%/4.00	77.8%/4.00
The questions can help prepare participants for the exercise	69.2%/3.77	38.5%/3.00		
I would use the scale routinely pre-exercise to help prepare participants for my exercise	38.5%/3.25	30.8%/2.92		
The scale allows to receive a detailed feedback/learn about participant' satisfaction with the exercise			77.8%/3.78	88.9%/4.25
I would use this scale for my exercise evaluation	46.21%/3.31	38.5%/3.33	44.4%/3.11	77.8%/4.00
The data obtained from this scale is useful for me	69.2%/3.62	53.8%/3.33	66.6%/3.67	88.9%/4.11
These questions are useful for participants		23.1%/2.75	33.3%/3.22	44.4%/3.22
It will be practical to use this scale in my exercise	30.8%/2.77	15.4%/2.50	33.3%/3.11	55.5%/3.44
These questions can be used for any exercise type	92.3%/4.00	45.2%/3.33	55.5%/3.11	77.8%/3.89
These questions can be used with any exercise participants	92.3%/4.00	38.5%/3.17	44.4%/3.33	66.7%/3.67
The best way to ask these questions is via an online survey	84.6%/4.15	69.2%/3.75	100%/4.33	77.7%/4.13

The experts were asked to complete the questionnaire and answer a set of usability questions separately for each of its four scales (Table 7 has the results) and provide their comments. Three scales (pre-exercise, exercise feedback and personal satisfaction with the exercise) were considered of an appropriate length, while the perceptions scale was perceived as lengthy. Even though most respondents agreed that the pre-exercise scale

could help prepare participants for the exercise (69.2%), only 38.5% would consider using it routinely to help with participants' preparation in their exercises. There was a strong agreement that pre-exercise scale could be used with any exercise (92.3%) and with participants in any emergency roles (92.3%). Experts agreed that the exercise feedback and personal satisfaction scales allow detailed feedback from participants on their experience; however, they also commented that the feedback scale is not universal, and is more applicable to tabletop exercises, mainly due to its detailed subscale of facilitator skills.

Fewer than half of the experts said they would consider using pre-exercise, perceptions and feedback scales for exercise evaluation in their own exercises, while 77.8% would consider using the personal satisfaction with the exercise scale for evaluation. Several comments provided by experts provided an insight into why they do not see much value in participants' individual responses. They are perceived as subjective opinions of people who "*come and go, [who] change jobs... I do not see it would add much*". Typically, exercise evaluation is supported by capturing group, not individual, responses by exercise evaluators to provide evidence for an exercise report: "*We are tasked to design the exercise, and evaluate the responses in order to produce an exercise report showing the lessons identified.*" Other barriers to routinely collecting survey data from exercise participants included perceived technical difficulties of collecting anonymous pre- and post-exercise data from participants, and a fear of collecting data that could potentially be damaging to organisational reputations: "*It is really helpful information academically and for the organisations but only if there is learning culture and people are open to criticism to identify gaps. If it was an internal exercise, the organisation could identify gaps in training and information dissemination and work on it but externally, you might not want to show your shortcomings*".

It was generally agreed that the best way to ask these questions is via an online survey.

4 Discussion

Collecting data from exercise participants can contribute to exercise evaluation, but this is not a routine practice (Skryabina et al., 2017), and usability analysis conducted as part of this study confirmed existing practices of relying on what exercise organisers perceive as more objective evidence for exercise evaluation. Evaluation data from participants, although recognised as important, was not considered useful for an exercise report by some EPRR experts.

One of the reasons for this attitude may be the lack of culture to use exercise participants' views to assess organisational performance. Another reason can be the lack of a valid and reliable tool to collect data from exercise participants. In this study, we report on the validity, reliability and usability of a four-scale questionnaire instrument that can be used to collect participants' reactions and perceptions from an emergency preparedness exercise, as well as to prepare participants for an exercise.

The effectiveness of HEPEs is difficult to determine, and one way of establishing it is to identify any immediate positive changes in participants' perceptions of having attended the exercise. The perception scale offers good reliability and sensitivity to change and allows measurement of participants' perceptions of their training and overall preparedness, organisational preparedness, preparedness for multi-agency response,

emergency plans, level of stress and competency-based knowledge and skills. Subscales of perceptions of team work and resources, which offer good internal consistency, had rather low test-retest agreement and therefore require further refinement. The perceptions scale is not context-specific and can be used in any exercise type.

The post-exercise feedback scale offers detailed feedback from participants on exercise conduct, by focusing on the major factors which affect learning from an exercise. The facilitator subscale was praised by experts for exploring in detail a variety of factors that contribute to good facilitation and enhanced learning outcomes from an exercise, but it has also been recognised that the subscale is more appropriate for a discussion-based exercise. However, all the other subscales of the post-exercise feedback scale (aims and objectives, exercise scenario, exercise format, key participants, exercise time, exercise evaluation and exercise venue) are not context specific and can be used with any exercise type and with any participants.

The pre-exercise scale offers a range of questions which can check participants' preparation, attitudes and expectations from an exercise. These can be of interest to exercise planners, and a majority of experts confirmed that these data would be useful to them. Literature suggests that pre-exercise assessment can facilitate participants preparation for an exercise, as reported with both discussion-based (Fowkes et al., 2010) and operation-based exercise participants (Kotora et al., 2014), by re-evaluating its value, purpose and personal objectives; this was also agreed by 69.2% of experts. However, only 30.8% of them considered that it was practical to use the scale in their exercises. Low test-retest agreement for this scale and for a few subscales on the exercise feedback scale may indicate the exercise specificity of the data. Almost all experts (92.3%) considered the pre-exercise scale applicable in any exercise context and with any exercise participants. Low test-retest agreement for question 47 ("I can locate my organisation's incident response plan") and the only significant difference observed in the responses of the control group participants provided two weeks apart on the 'perceptions of emergency plans' subscale may indicate that taking part in the survey prompts actions that are under participants control, like improving the knowledge of emergency plans.

The personal satisfaction scale provides the data for an overall measure of participants' satisfaction with the exercise, as well as their behavioural intentions and capacity to engage in actions to improve their individual and organisational preparedness in response to the lessons identified in the exercise. The data from this scale was perceived as the most useful to practitioners (88.9%) and 77.8% were prepared to use the scale as an exercise evaluation tool. Psychometric properties for this scale were good both for discussion-based and operation-based exercises, and it can be applied in any exercise context and with any participants.

The questionnaire can be offered as an online tool, which can be completed by participants pre- and post-exercise as a part of exercise evaluation for any exercise type and with any participants. A coding system can be offered as an option to link pre- and post-exercise responses while preserving participants' anonymity.

Usability analysis clearly indicated that collecting views of participants is not perceived as a common practice, in line with our previous in-depth analysis of recently published literature in the field (Skryabina et al., 2017). The least useful scale for practitioners' interests appeared to be the perceptions scale; the scale which allows participants' perceptions of their own and their organisational level of emergency preparedness to be studied. A few barriers identified from the usability study include the lack of trust in participants' individual views of emergency preparedness and the lack of

learning culture of involved organisations, which are consistent with the previously reported barriers to organisational learning (Coles, 2014). We acknowledge, however, that the views collected may just be a peculiarity of the group of experts who took part in the usability study. A wider range of views could be sought in future work, including views of exercise participants, and more work to develop the culture and opportunities for individual participant development as part of exercises.

Exercises are important learning opportunities for individual staff and routinely collecting data from participants can not only enhance the quality of the exercise evaluation, but can also serve an educational purpose for participants by promoting self-reflection and self- evaluation (Skryabina et al., 2018) as well as play a role in motivating and prompting post-exercise actions (Fowkes et al., 2010). If we can characterise and study the impact that an exercise is having on individuals, we have an opportunity to design better exercises, which are more effective and are in turn better value for money.

5 Limitations

Despite efforts to increase the response rate for the test-retest reliability study from exercise participants, only 15 (20%) out of 73 individually invited exercise participants completed the first questionnaire and only nine (12%) completed the second. No opportunity was available to engage operation-based exercises in the test-retest study. Low test-retest agreement for some pre-exercise scales and the exercise feedback scale may indicate the exercise specificity of the data and a larger sample for the test-retest study would be useful to clarify this. The questionnaire could not differentiate perceptions of less experienced staff (fewer than four years of experience) from perceptions of more experienced staff (more than nine years of experience) for operation-based MTN exercises, although some differences were identified for discussion-based exercises. The reason may be in the nature of the operation-based exercises, which tend to focus more on ‘training’ and thus involve more a homogeneous group of participants who need training (Leiba et al., 2006; Motola et al., 2015; Summerhill et al., 2008), while discussion-based exercises tend to involve participants with a wider range of experiences to facilitate discussions (Adini et al., 2015; Gin et al., 2013; Alison et al., 2015). However, it also needs to be acknowledged that the sample size of the more experienced group in MTN exercises was low ($N = 14$), which might have also affected the outcomes. The questionnaire was considered to be rather lengthy by EPRR experts, particularly the perceptions scale, and it may not be practical to use all its subscales with every participant in every exercise. However, the subscales cover a variety of variables that contribute to emergency preparedness competencies, and which are typically addressed in emergency preparedness exercises as part of emergency preparedness. The users have flexibility to use only those subscales they perceive appropriate for their needs. Although an invitation to comment on questionnaire construction and usability was offered to all exercise participants who took part in the test-retest study, no comments were provided. The facilitator subscale of the exercises feedback scale was designed specifically for discussion-based exercises and takes into account elements of group-discussion facilitation. Although facilitation is an important element of operation-based exercises, the evidence on the best ways of providing it is limited for this type of emergency preparedness exercises.

Further work will identify factors contributing to effective facilitation of operation-based exercises. Evidence of the optimal timing for the pre-exercise assessment is lacking and further work will be advocated to determine how far in advance pre-exercise evaluation should be conducted to enable participants to suitably prepare, and how far in advance is optimal to focus participants attention to facilitate their engagement with the exercise.

The overall aim of HEPEs is to enhance preparedness but the literature on retention and transfer of learning is sparse (Skryabina et al., 2017). It is also reported that there is a tendency to identify the same lessons repeatedly suggesting knowledge gained does not transfer to practice (Coles, 2014). Evaluation methods that are limited to pre- and post-measures to identify for any change are not suited to evaluating the retention and transfer of knowledge. Therefore, future research should seek to develop follow-up measures that evaluate the impact of the exercise in the months following exercise participation.

6 Conclusion

The questionnaire provides a reliable and valid tool for exploring healthcare staff experiences of participating in an emergency preparedness exercise, and their perceptions of the impact of the exercise on their own and their organisation's emergency preparedness. The four scales of the questionnaire can provide detailed feedback on participants' attitudes, perceptions, exercise conduct, learning and post-exercise intentions. The pre-exercise assessment scale provides data on the participants' attitudes towards an exercise, and also serves as an exercise preparation tool. The questionnaire construct allows using different scales in different combinations dictated by practical needs. However, usability analysis confirmed existing practices of relying on evidence collected from expert evaluators for exercise evaluation and the lack of culture to include participants' in this process. The practicality to use a participant questionnaire was one of the identified barriers to its use by professionals delivering exercises, and consideration of extra resources at the exercise planning stage may need to be in place to promote participants evaluation at exercises. Other barriers were associated with exercise scope, and the applicability of the questionnaire to organisational-level exercise objectives; individual level objectives were missing from the agenda. Offering a valid and reliable tool to collect data from exercise participants can not only help with the overall exercise evaluation and provide feedback to exercise planners on the exercise conduct, but can also enhance the educational value of the exercise to participants. This may have positive implications to overall emergency preparedness through better prepared staff as well as organisations.

Acknowledgements

This study was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emergency Preparedness and Response, a partnership between Public Health England, King's College London and the University of East Anglia. The views expressed are those of the authors and not necessarily those of the NIHR, Public Health England or the Department of Health and Social Care. Professor

Richard Amlôt is also supported by the NIHR HPRUs in Evaluation of Interventions at Bristol University and Modelling Methodology at Imperial College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or Public Health England.

Funding

This study was funded by the National Institute for Health Research Health Protection Research Units (HPRU-2012-10414);

Conflict of interest

The authors declare that there is no conflict of interest

References

Adini, B., Aharonson-Daniel, L. and Israeli, A. (2015) 'Load index model: an advanced tool to support decision making during mass-casualty incidents', *The Journal of Trauma and Acute Care Surgery*, Vol. 78, pp.622–627.

Alison, L., Power, N., Van Den Heuvel, C., Humann, M., Palasinksi, M. and Crego, J. (2015) 'Decision inertia: deciding between least worst outcomes in emergency responses to disasters', *Journal of Occupational and Organizational Psychology*, Vol. 88, pp.295–321.

Anderson, C. (1998) 'On the scene. countdown to disaster: how to plan an effective disaster drill', *Emergency Medical Services*, Vol. 27, pp.59–61.

Beaton, R.D., Oberle, M.W., Wicklund, J., Stevermer, A., Boase, J. and Owens, D. (2003) 'Evaluation of the washington state national pharmaceutical stockpile dispensing exercise: Part I--patient volunteer findings', *Journal of Public Health Management and Practice: JPHMP*, Vol. 9, pp.368–376.

Beerens, R.J.J. and Tehler, H. (2016) 'Scoping the field of disaster exercise evaluation – A literature overview and analysis', *International Journal of Disaster Risk Reduction*, Vol. 19, pp.413–446.

Biddinger, P.D., Savoia, E., Massin-Short, S.B., Preston, J. and Stoto, M.A. (2010) 'Public health emergency preparedness exercises: lessons learned', *Public Health Reports*, pp.100–106.

Coles, E. (2014) 'Learning the lessons from major incidents: a short review of the literature', *Emergency Planning College*, Vol. 10, pp.1–22.

ECDC (2014) *Handbook on Simulation Exercises in Public, EU Health Settings*, European Centre for Desease Prevention and Control (ECDC), https://ecdc.europa.eu/en/publications-data/handbook-simulation-exercises-eu-public-health-settings

Emery, R.J., Sprau, D.D., Morecook, R.C. and Herbold, J. (2009) 'Surge capacity volunteer perspectives on a field training exercise specifically designed to emphasize likely roles during a disaster response', *Health Physics*, Vol. 97, pp.S155–S160.

ETS (2016) *Emergo Train System* (*ETS*) [Online], https://www.emergotrain.com/ [Accessed].

Evaluation of Exercises Handbook (2011) *Evaluation of Exercises Handbook*, Swedish Civil Contingencies Agency.

Ford, J.K. and Schmidt, A.M. (2000) 'Emergency response training: strategies for enhancing real-world performance', *Journal of Hazardous Materials*, Vol. 75, pp.195–215.

Fowkes, V., Blossom, H.J., Sandrock, C., Mitchell, B. and Brandstein, K. (2010) 'Exercises in emergency preparedness for health professionals in community clinics', *Journal of Community Health*, Vol. 35, pp.512–518.

Freimuth, V.S., Hilyard, K.M., Barge, J.K. and Sokler, L.A. (2008) 'Action, not talk: a simulation of risk communication during the first hours of a pandemic', *Health Promotion Practice*, Vol. 9, pp.35S–44S.

Gin, J.L., Chan, E.W., Brewster, P., Mitchell, M.N., Ricci, K.A., Afable, M.K. and Dobalian, A. (2013) 'Using exercises to identify veterans health administration priorities for disaster response: findings from the new madrid earthquake training exercise', *J. Public Health Manag Pract*, Vol. 19, pp.126–132.

Henning, K.J., Brennan, P.J., Hoegg, C., O'Rourke, E., Dyer, B.D. and Grace, T.L. (2004) 'Health system preparedness for bioterrorism: bringing the tabletop to the hospital', *Infection Control and Hospital Epidemiology*, Vol. 25, pp.146–155.

High, E.H., Lovelace, K.A., Gansneder, B.M., Strack, R.W., Callahan, B. and Benson, P. (2010) 'Promoting community preparedness: lessons learned from the implementation of a chemical disaster tabletop exercise', *Health Promotion Practice*, Vol. 11, pp.310–319.

Johnstone, A.H., Sleet, R.J. and Vianna, J.F. (1994) 'An information processing model of learning: its application to an undergraduate laboratory course in chemistry', *Studies in Higher Education*, Vol. 19, pp.77–87.

Kirkpatrick, D. (1998) *Evaluatinig Training Programs: The Four Levels*, Berrett-Koehler, New York.

Kotora, J.G., Clancy, T., Manzon, L., Malik, V., Louden, R.J. and Merlin, M.A. (2014) 'Active shooter in the emergency department: a scenario-based training approach for healthcare workers', *American Journal of Disaster Medicine*, Vol. 9, pp.39–51.

Leiba, A., Goldberg, A., Hourvitz, A., Amsalem, Y., Aran, A., Weiss, G., Leiba, R., Yehezkelli, Y., Goldberg, A., Levi, Y. and Bar-Dayan, Y. (2006) 'Lessons learned from clinical anthrax drills: evaluation of knowledge and preparedness for a bioterrorist threat in Israeli emergency departments', *Annals of Emergency Medicine*, Vol. 48, pp.194–199.

Leung, S. (2011) 'A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point likert scales', *Journal of Social Service Research*, Vol. 37, pp.412–421.

Morris Jr., J.G., Greenspan, A., Howell, K., Gargano, L.M., Mitchell, J., Jones, J.L., Potter, M., Isakov, A., Woods, C. and Hughes, J.M. (2012) 'Southeastern center for emerging biologic threats tabletop exercise: foodborne toxoplasmosis outbreak on college campuses', *Biosecurity and Bioterrorism*, Vol. 10, pp.89–97.

Motola, I., Burns, W.A., Brotons, A.A., Withum, K.F., Rodriguez, R.D., Hernandez, S., Rivera, H.F., Issenberg, S.B. and Schulman, C.I. (2015) 'Just-in-time learning is effective in helping first responders manage weapons of mass destruction events', *J. Trauma Acute Care Surg.*, Vol. 79, S152-6.

Perry, R.W. (2004) 'Disaster exercise outcomes for professional emergency personnel and citizen volunteers', *Journal of Contingencies and Crisis Management*, Vol. 12, pp.64–75.

Peterson, D.M. and Perry, R.W. (1999) 'The impact of disaster exercise on participants', *Disaster Prevention and Management*, Vol. 8, pp.241–254.

Petty, R.E. and Cacioppo, J.T. (1981) *Attitude and Persuation: Classic and Contemporary Approaches*, C. Brown Company.

Savoia, E., Biddinger, P.D., Fox, P., Levin, D.E., Stone, L. and Stoto, M.A. (2009a) 'Impact of tabletop exercises on participants' knowledge of and confidence in legal authorities for infectious disease emergencies', *Disaster Medicine and Public Health Preparedness*, Vol. 3, pp.104–110.

Savoia, E., Testa, M.A., Biddinger, P.D., Cadigan, R.O., Koh, H., Campbell, P. and Stoto, M.A. (2009b) 'Assessing public health capabilities during emergency preparedness tabletop exercises: reliability and validity of a measurement tool', *Public Health Reports*, Vol. 124, pp.138–148.

Savoia, E., Biddinger, P.D., Burstein, J. and Stoto, M.A. (2010) 'Inter-agency communication and operations capabilities during a hospital functional exercise: reliability and validity of a measurement tool', *Prehospital and Disaster Medicine*, Vol. 25, pp.52–58.

Savoia, E., Agboola, F. and Biddinger, P.D. (2012) 'Use of after action reports (AARs) to promote organizational and systems learning in emergency preparedness', *International Journal of Environmental Research and Public Health*, Vol. 9, No. 8, pp.2949–2963.

Skryabina, E., Reedy, G., Amlôt, R., Jaye, P. and Riley, P. (2017) 'What is the value of health emergency preparedness exercises? A scoping review study', *International Journal of Disaster Risk Reduction*, Vol. 21, pp.274–283.

Skryabina, E., Riley, P., Reedy, G. and Amlôt, R. (2018) 'A scoping review of evaluation methods for health emergency preparedness exercises', *American Journal of Disaster Medicine*, Vol. 13, pp.107–127.

Stergachis, A., Wetmore, C.M., Pennylegion, M., Beaton, R.D., Karras, B.T., Webb, D., Young, D. and Loehr, M. (2007) 'Evaluation of a mass dispensing exercise in a cities readiness initiative setting', *American Journal of Health-System Pharmacy*, Vol. 64, pp.285–93.

Summerhill, E.M., Mathew, M.C., Stipho, S., Artenstein, A.W., Jagminas, L., Russo-Magno, P.M., Potter, S. and Shapiro, M.J. (2008) 'A simulation-based biodefense and disaster preparedness curriculum for internal medicine residents', *Med Teach*, Vol. 30, pp.e145–151.

Williams, J., Nocera, M. and Casteel, C. (2008) 'The effectiveness of disaster training for health care workers: a systematic review', *Annals of Emergency Medicine*, Vol. 52, pp.211–222.e2.