safety

MDPI

Article

Developing a Time-Based Evaluation Method for Functional Exercises of Emergency Medical Operations

Ngoc Thang To [1], Takaaki Kato [2,*], Misaki Imai [2,3], Masao Taninobu [4], Kazuaki Kohriyama [5] and Shigehiko Ito [6]

1 Graduate School of Environmental Engineering, The University of Kitakyushu, Kitakyushu 808-0135, Japan
2 Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu 808-0135, Japan
3 ICON Japan K. K., Tokyo 105-0001, Japan
4 Kitakyushu Fire Department, Kitakyushu 803-8509, Japan
5 Foundation for Ambulance Service Development, Kitakyushu 807-0874, Japan
6 Kitakyushu City Yahata Hospital, Kitakyushu 805-8534, Japan
* Correspondence: tkato@kitakyu-u.ac.jp; Tel.: +81-93-695-3237

Received: 22 May 2019; Accepted: 6 August 2019; Published: 7 August 2019

check for updates

Abstract: Public health service is one of the most important sectors in terms of saving lives. During a disaster, hospitals and medical groups implement extension tasks from their daily activities. Enhancing coordination across organizations contributes to the removal of communication barriers. Functional exercises are simulated trainings for emergency responders that aim to enhance coordination capabilities. The application of time elements in exercise evaluation methods is a significant area of potential research. We develop methods to quantitatively analyze time spent on completing unit tasks in functional exercises. This study focuses on analyzing observed time data in two functional exercises of the Disaster Medical Operation Center in Kitakyushu, which were repeated in October and November 2015. We employed a censored regression method to analyze the time spent on both complete and incomplete unit tasks together. Differences in processing time for 39 tasks, which were repeated in the two exercises, are visually inspected. Benefits of time study in the evaluation of exercises are presented.

Keywords: disaster preparedness; functional exercise; evaluation; task processing time

1. Introduction

Disaster management, which is also referred to as emergency management, is defined as the profession that addresses the management of disasters [1–3]. Emergency medical services are one of the primary components of disaster management for saving human lives [4–6]. Since many different organizations need to work together to provide emergency medical services, designing exercises to identify problems of communication, coordination, and cooperation among these organizations is necessary [7]. Moreover, since major incidents for emergency medical services do not occur frequently, organizations and staff need to practice the procedures and skills before these events to have better preparedness [8,9]. Emergency preparedness activities include a complex cycle of planning, equipment, training, practice, and improvement [10,11], while emergency exercises have been identified as a key component of the cycle [12,13]. Training and exercise are also defined as specific nonmaterial activities to reduce the exposure risk component [14]. Education and training related to disaster can enhance the preparedness of individuals that contribute to risk reduction [15].

Emergency exercises can be classified into two major groups including discussion and operation-based exercises. The standard training program mentioned in the Homeland Security

Exercise and Evaluation Program (HSEEP) by the US Department of Homeland Security includes seven types of exercises, according to their block-building approach [16]. Discussion-based exercises include seminars, workshops, tabletop activities, and games. Operation-based exercises include drill, functional, and full-scale exercises. In the National Incident Management System (NIMS) of the US, a well-planned exercise program scheme to improve integration and interoperability should include orientation and a tabletop exercise, followed by a functional exercise [17]. Among these, functional exercises are most suitable for improving interagency coordination, clarifying roles and responsibilities, and identifying problems in emergency management plans [18,19]. In a functional exercise, a series of simulated emergency events, or "exercise injects", are provided to exercise participants, or "players", and together, they find solutions to the problems specified in these injects [20]. There are a number of injects delivered in a functional exercise that make the event appear real. Players order a resource but do not actually deploy them, unlike in drills. For public health emergencies, functional exercises are simulated by providing health-centered injects to players. One of the main benefits of a functional exercise is to observe how participants respond to reality-based injects [21]. This type of exercise also helps to evaluate the current capability of medical stakeholders in emergencies, which is difficult to observe in detail in an actual disaster [22].

Methods and tools to evaluate functional exercises for better disaster preparedness are still under development. As one type of operation-based group, evaluation of functional exercises focuses on effectiveness of communication and adequacy of cooperation among stakeholders [16,23]. Through exercise, it also verifies the preparedness of organizations before a real emergency occurs [24]. Russo et al. [14] developed an approach to measure risk reduction through training activities. In emergency medical operations, an unevaluated or inadequately evaluated plan and exercise may do more harm because it leads to underestimated risk and poor performance in actual emergencies [23]. There are two main evaluation criteria in a functional exercise: how well the exercise result met its objectives, and how the players performed their tasks [25]. This study is concerned with the second criterion, which is used to improve emergency management plans and the design of future exercises. Some tools have recently been developed for evaluating performance during exercises. The HSEEP [16] provides a template, the Exercise Evaluation Guide (EEG), as a consistent tool to evaluate trainees' performance and rate their achievements in a target rating system. Another evaluation instrument for public health emergency preparedness exercises, developed by the Harvard School of Public Health (HSPH) [26], consists of a combination of action checklists, subjective scoring, and subjective comments. Savoia et al. [22] list common instruments to measure performance during exercises, including checklists, scores, and open-ended questions. They also highlight the necessity of gathering data in quantitative or standardized ways rather than in narrative form in public health system exercises. A review paper by Skryabina et al. [8] reveals that most studies on health emergency preparedness exercises use qualitative data from self-reporting and participants' perceptions of exercise outcomes. Nelson et al. [27] argued that there is a lack of metrics involving time-based attributes to observe and evaluate exercises. Russo et al. [24] noted that the principal aim of exercise and training is to reduce the response time in an actual emergency. Moreover, the number of lives saved is proportional to the speed and efficiency of emergency medical operations [4,28,29]. Therefore, time is an important factor that should be given more consideration when evaluating performance during medical emergency exercises.

There are limitations in the present use of time to evaluate performance in functional exercises. The template of both the EEG and the evaluation instrument of the HSPH for health emergency preparedness exercise programs do not have clear sections for measuring task performing time. There are some evaluation forms in existing functional exercises that include a time element. In particular, under the guidance of the Federal Emergency Management Agency in the US, functional exercises of the Community Emergency Response Team (CERT) apply an observation form that records the completion time for each exercise inject. Some exercise evaluation guidance manuals published by the California Department of Fish and Game [30] and the International Atomic Energy Agency [12] include a timeline in the observation form. The Public Health Emergency Exercise Toolkit [31], which provides a timesheet for recording "time in" and "time out" at stations involved in a full-scale public health exercise, was developed at Columbia University. A recent development of functional exercises is the use of IT technology to record communications among players. For example, WebEOC [32], a product by Juvare, LLC is used in many emergency operation centers (EOCs) to provide support for resources and task management. Time of players' action, communication partners, and descriptions of the activities are recorded in the incident notification log via WebEOC. However, time study of task-time has not been integrated into the modules of this software. Therefore, a serious consideration of time elements can contribute to further advancement in this field. As we will demonstrate, time can provide more information about the performance of exercise players.

This study focuses on the evaluation program of functional exercises in emergency medical operations. Since time is an important element to evaluate participants' performance, we aimed at developing a method to analyze the time required to complete emergency management tasks in functional exercises. To realize this, we incorporated the knowledge of time studies in the industrial domain, communication studies, and team performance studies. Results are expected to provide clues to improve job performance of the emergency management organization.

In the next section, we outline the data used in this study and explain a method to analyze the results of time measurement. The study was based on data of the two medical functional exercises conducted in October and November 2015 by the Disaster Medical Operation Center (DMOC) in Kitakyushu, Japan. Section 3 provides our results and discussion. Section 4 draws some conclusions and outlines future implications for developing evaluation methods of functional exercises.

2. Data and Methodology

2.1. Functional Exercise for DMOC

Kitakyushu is an industrial city with a population of one million inhabitants that borders the northern region of Kyushu Island in Japan. The main types of natural hazards that threaten Kitakyushu include floods, typhoons, landslides, and earthquakes. Although the city has not suffered from major disasters in the last 50 years, several major disasters occurred in the areas surrounding the city, including earthquakes in Kumamoto in 2016 and heavy rain in Northern Kyushu in 2017. Disaster preparedness is necessary for the city to reduce damage during disasters, as well as to support other cities.

The Kitakyushu Medical Association (KMA) and the local government have been developing a medical communication and resource management system for emergency situations. In 2016, the KMA's Emergency Medical Plan (EMP) provided a framework for communication among medical stakeholders in Kitakyushu. The DMOC was established to coordinate medical organizations including hospitals and the KMA to meet various medical needs during disasters. Under supervision by the city's EOC, the DMOC cooperates with other stakeholders to allocate appropriate medical resources. When a disaster occurs, six groups of the DMOC gather at the Kitakyushu City Yahata Hospital, which is the city's disaster base hospital. There is one general coordination group and five specialized groups including "Site", "Base Hospital", "Shelter", "Logistic", and "Reinforcement" groups. "Site", "Base Hospital", and "Shelter" groups are in charge of communicating with corresponding organizations in these areas. The "Logistic" group distributes equipment and goods to shelters, medical facilities, and disaster sites. The "Reinforcement" group responds to disaster response support offers from outside organizations. The leader of the DMOC is the director of the Disaster Medical Training Center of Kitakyushu City Yahata hospital. Members of the DMOC are from the KMA, city government, and other medical organizations.

Functional exercises are specified in the EMP as an important program to improve the awareness of related personnel and enhance coordination mechanisms among stakeholders in Kitakyushu. Through experiencing unusual situations in simulated exercises, participants can improve their readiness for an emergency as well as review and update existing plans and operations [33].

On October 27 and November 21 in 2015, two functional exercises were conducted in Kitakyushu City Yahata Hospital to simulate medical emergencies. These exercises simulated damages to the city caused by an earthquake with a magnitude of 6.9 on the Richter scale. The assumed earthquake and its damage were the same between the two exercises. Exercise time was for one and a half hours, and each exercise consisted of two sections. Section 1 simulated the situation immediately after the earthquake, and the initial set up of the DMOC was the major activity during this section. Section 2 simulated the situation in which the DMOC coordinated medical activities in evacuation shelters and other locations. In DMOC exercises, the player team was separated into five groups: "Coordination", "Site", "Base Hospital", "Shelter", and "Logistics and Reinforcement" groups. These groups were in charge of responding to injects delivered from the controller team via telephone. Most players joined both the exercises.

2.2. Work Breakdown and Unit Task-Time Measurement

Developing a work breakdown structure (WBS) is a common approach to analyze complicated work and identify unit tasks that are required to complete the work. WBS has been applied for planning of emergency management [34–36]. The Kitakyushu Fire Department (2008) [37] identified workflow to respond to each exercise inject by specifying relevant unit tasks in the WBS and the network structure of communications among emergency management groups that were responsible for processing these unit tasks. They then measured the time required to complete each unit task and summarized the measured time to identify strengths and weaknesses of the emergency management activities regarding each emergency support function. Figure 1 shows an example of the workflow and time measurement result created from our case study of the inject code 19 of the October DMOC exercise. We refer to this workflow as the "communication network". We followed this unit time measurement approach and developed methods to analyze the time record in detail and compare the two time records obtained from repeated functional exercises.

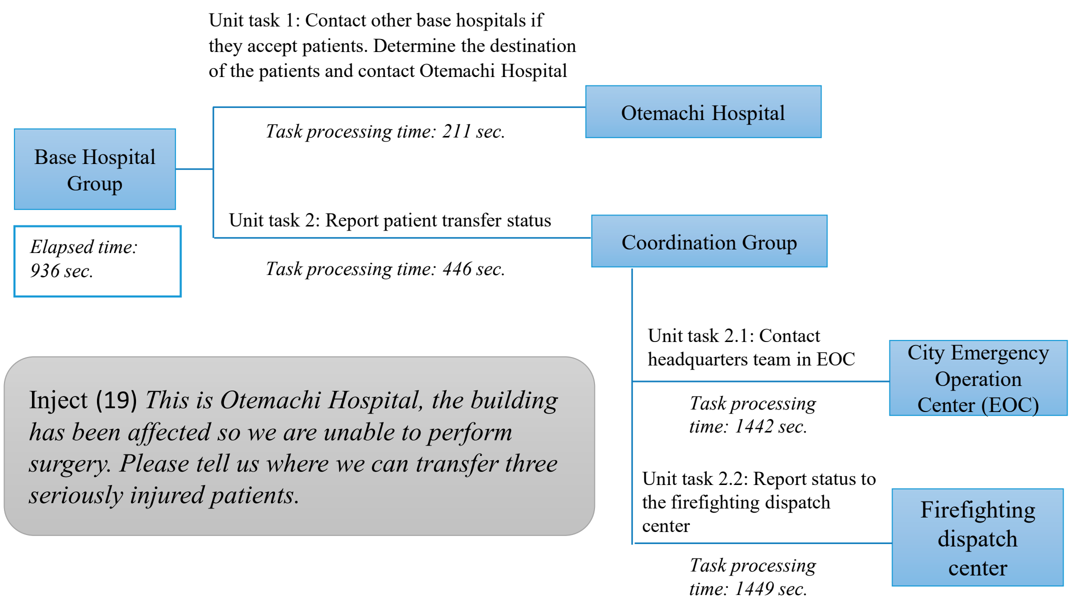


Figure 1. Communication network for inject code 19 of the October Disaster Medical Operation Center (DMOC) exercise.

We defined several measures of time. Start time of a unit task was defined as the time when a group began to act after it received information from the preceding group in the communication network. End time was defined as the time when the group completed the unit task. Task processing time was the difference between the end time and the start time. Elapsed time was defined as the length of time between the start time of a unit task and the beginning of the exercise. Start and end times of each unit task were recorded by evaluators using an exercise evaluation software developed by Infogram Inc. Note that the same group may have conducted some different unit tasks simultaneously to respond to the information sent to the group; these tasks had the same elapsed time. We collected the results of the task processing time in both the October and November exercises. There were 17 exercise emergency injects in October and 18 in November. DMOC groups needed to complete more than one unit task to find a solution to an exercise inject. Thus, a total of 57 unit tasks and 59 unit tasks were conducted in October and November, respectively. These two exercises had the same 39 unit tasks. Three unit tasks in October and five unit tasks in November remained incomplete until the end of the exercise. We were unable to ascertain the end time and task processing time for these unit tasks.

2.3. Defining Stages in the Communication Network

There are a variety of time analysis techniques in industrial engineering and other fields. Time study has evolved from the basic concept introduced by Taylor [38], who proposed the use of performance standards for tasks in factories based on his study on processing time. At the basic level of time studies, job and work are broken down into unit tasks, and the time consumed for completing each unit task is then measured [39]. In the Toyota Production System, the completion time per unit of output is measured in each unit task that is used to construct standard operating routines. In this procedure, dividing the lead time of production into the lead time of data processing, that of the manufacturing activity itself, and that of delivering completed products to customers help identify causes of wasted time [40]. We applied this concept to our analysis of the communication network, and divided one communication network into two stages. For example, the unit tasks assigned to the "Base Hospital" group in Figure 1 were in Stage 1. The unit tasks assigned to the "Coordination" group were in Stage 2.

2.4. Defining Unit Task Types

We defined unit task types to compare task processing time across those types. We referred to communication studies to classify unit tasks. First, in work and organizational psychology, an integrated method of task analysis is hierarchical task analysis (HTA), which was introduced by Annett and Duncan [41]. One of the HTA categories used for analyzing teamwork is "communication", and three observable actions in this category are "Send information", "Receive information", and "Discuss" [42]. Second, Habermas [43] distinguishes "task-oriented" and "relationship-building" as two types of communication among groups. The first category refers to when a speaker contacts a listener to request certain actions toward goals, while the second category is oriented toward sharing common information and understanding. We examined group tasks in the DMOC exercises and created four task types that were used for statistical analyses in the next section. The four task types are:

- "Request" refers to actions that send information to other groups to request certain actions toward goals. This is to send information in a task-oriented manner.
- "Report" refers to when a group sends information to others to share information and understanding. This is to send information in a relationship-building manner.
- "Inquiry" is defined as when one group collects or receives information from other groups to conduct subsequent tasks.
- "Decision making" is defined as a set of actions in one group that includes discussion, making decisions, and delegating decisions to other groups. This unit task includes making discussions among relevant players.

2.5. Team Performance Factors

There are team performance factors that should affect the task processing time. Often-cited models of team performance for meeting targets include those of McGrath, Gladstein, and Hackman [44–46]. All these models follow the basic concept of the input–process–output sequence. Although each model has some unique additional factors, there are common factors included in these models. Three input factors including individual, team, and environmental levels found in McGrath's model were considered as potential factors regarding task processing time in the DMOC exercises. The temporal phase of the exercise was added as another potential factor [47]. Elapsed time and stage in communication network were used as variables to proxy this temporal phase factor.

Additional factors may affect the performance of emergency responders. They often work under stressful conditions that need long working hours and need difficult tasks with incomplete information to be done in a short time [48,49]. Occupational stressors for emergency responders include working environment, time pressures, and workload, which are listed in the Report of the National Institute of Mental Health [50]. Effects of stress on team members before, during, and after emergency events may cause problems physically (e.g., fatigue, headache), behaviorally (e.g., violence, alcohol abuse) and emotionally (e.g., fear, anxiety) [50,51]. During an emergency response, the level of environmental stress affects team performance in several ways. Stressful condition may cause problems in the decision-making processes of teams [52] or conflicts among team members [53]. Heavy workload or time pressure can cause ineffective team performance [54]. However, it is possible that an adequate level of stress increases situational awareness of team members and improves their performance [49,55]. DMOC is an emergency operation center (EOC). EOC members face with some stressors, such as high workloads in exceptionally long working hours, pressure to achieve certain goals, and unfamiliar work environments [56,57]. We considered workloads and working hours as potential stressors in DMOC exercises. Workloads were measured by the number of simultaneously processed unit tasks by a DMOC group. Effect of working hours was considered by examining the relationship between elapsed time of each exercise inject and unit task processing time.

Table 1 summarizes the potential factors affecting task processing time and the availability of data obtained from the DMOC exercises.

Table 1. Factors affecting task processing time.

Potential Factors of Task Processing Time		Availability of Data in DMOC Exercises
Individual level	Member skills and training	*Not observed*
	Personality characteristics of team member	*Not observed*
Group level	Structure of group	Communication domain
	Level of cohesiveness	*Not observed*
	Group size	Number of members in each group
Environmental level	Task characteristic	Unit task types
	Level of heavy workload at a particular time	Number of simultaneously processed unit tasks in a group
Temporal phase	Stages of team project	Elapsed time Stage in communication network

2.6. Statistical Analysis

We first visually examined task processing time for each exercise using a box plot. We applied both a parametric (ANOVA) and a nonparametric method (Kruskal–Wallis test) to compare the mean and median, respectively, of task processing time across the five DMOC groups and four task types. The hypothesis of ANOVA test is "H_0: there is no difference among means of task processing time". If the p-value of the test is smaller than 0.05, the test rejects H_0. In this case, we used the Tukey's test to determine the pairs of means that were statistically different from each other. We then used a censored regression model to distinguish the impacts of multiple independent variables on task processing time in each exercise. The processing time for the unfinished tasks in the exercises were considered as right-censored data in the censored regression model and contributed to utilizing the maximum available sample size for the analysis. Maddala [58] provides further information on censored regression models. Table 2 shows a list of the independent variables used for our censored regression analysis.

Table 2. Independent variables in multiple regression model.

Variable	Variable Types (Unit)
DMOC group	Dummy variable Coordination Site Base Hospital Shelter Logistics and Reinforcement (L&R)
Communication domain	Dummy variable Within DMOC With outside groups
Unit task type	Dummy variable Report Request Inquiry Decision making
Simultaneously processed unit tasks	Countable number (piece)
Elapsed time	Continues number (second)
Stage in communication network	Dummy variable Stage 1 Stage 2

Because some variables were not observed in the DMOC exercises, the "DMOC" group variable reflected several dimensions of each group, such as member skills, member characteristics, internal group structure, and group cohesiveness. Since the number of group members was perfectly correlated with the dummy variables of the "DMOC" group, this variable was excluded from the analysis to avoid multicollinearity. After conducting these analyses for each exercise, we visually inspected the differences in processing time for the tasks that were examined both in the October and November exercises.

3. Results and Discussion

3.1. Visualization of Whole Communication Network Spanning across Groups

Before proceeding to our analysis of task processing time, we provide an outline of the entire network of communication across DMOC groups. Figure 2 shows the communication networks that synthesized communication networks created for each exercise inject. These two charts show that the patterns of the communication links among the player groups were the same between the October and November exercises, reflecting the fact that the majority of the exercise injects were identical between these exercises. The "Coordination" group worked as a hub of communication among player groups and processed the largest number of unit tasks in both exercises.

There were a high number of communication links between the "Coordination" group and "Other organization" to respond to simulated injects in these exercises. These frequent communications were related to risk reduction by reducing society's exposure in an emergency. In medical emergency operations, it is necessary to provide timely mass-patient care services to shelters for disaster victims, which helps reduce exposure to human health hazards. Since the number of victims in a disaster may increase suddenly and exceed the response capacity of an individual medical organization, more resources from other organizations should be used. Therefore, communications between the "Coordination" group and other organizations were mainly related to timely sharing of information and updating the current emergency response processes. These tasks aimed to improve the readiness of using larger resources from other organizations for mass-patient care services. Moreover, major responses in emergency medical services happen quite infrequently [8]; there is less chance of communication between the DMOC "Coordination" group and other organizations in daily activities.

Frequent communications in exercises help to reduce risks of communication failure among groups. These training activities contributed to reduce exposure in an actual emergency.

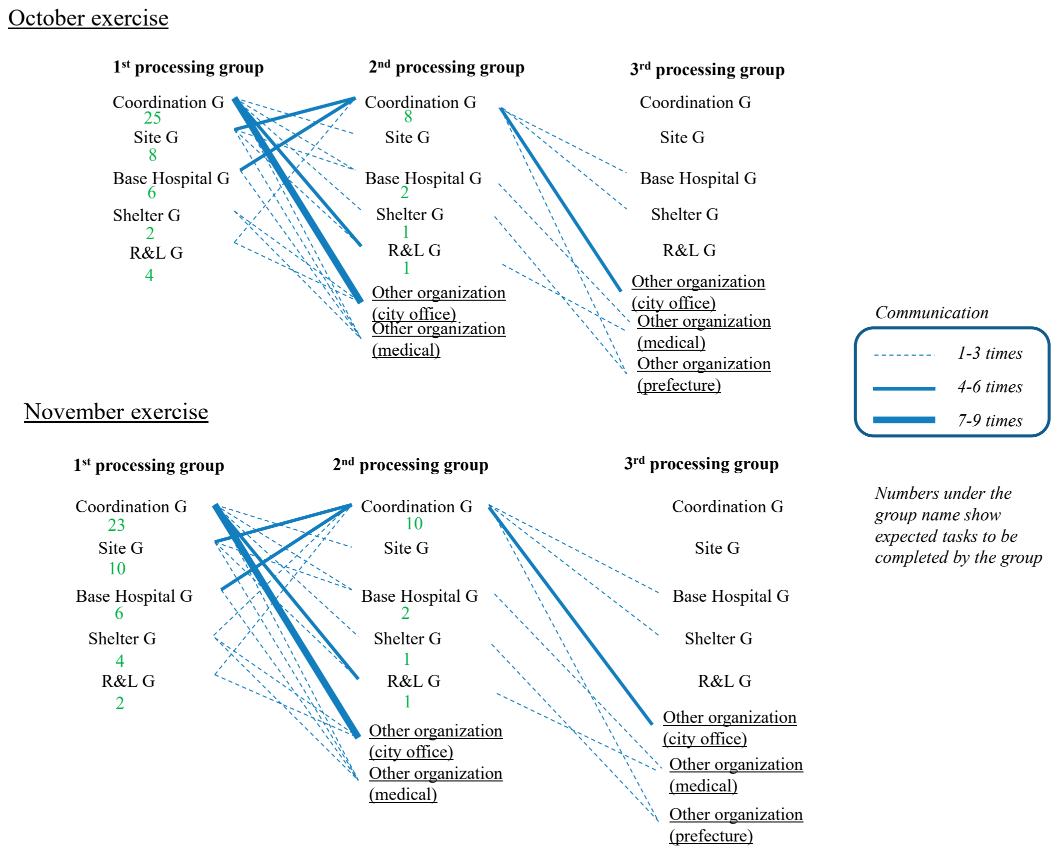
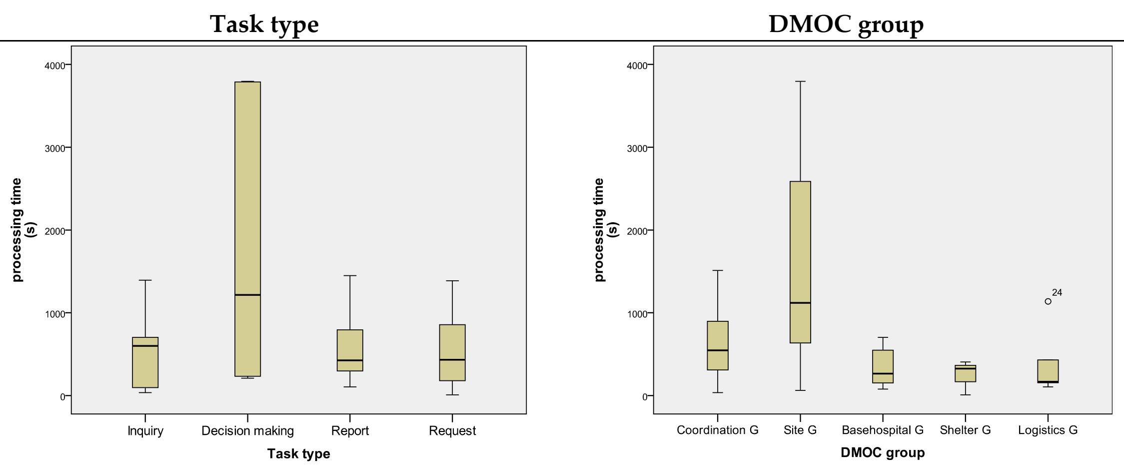


Figure 2. Communication networks of the two exercises.

3.2. *Factors of Task Processing Time*

3.2.1. Analysis of the October Exercise

We first present the results from the October exercise. Box plots are used to show the distribution of task processing time among five groups and four types of unit task in this exercise. The minimum, first quartile, median, third quartile, and maximum value of the data are displayed. Figure 3 summarizes the task processing time for unit task types and DMOC groups. Since the p-value of ANOVA for task processing time was smaller than 0.05, the Tukey's test was applied to compare mean values among unit task types and types of group. Then, at the 5% significance level, "decision making" took more time than the other three unit task types, and the "Site" group took more time to process their unit tasks than the other four groups.



Task type	N	Mean (s)	Std. Deviation (s)	Min (s)	Max (s)
Inquiry	9	528.00	448.295	36	1395
Decision making	6	1743.67	1658.311	211	3796
Report	22	581.05	417.021	105	1449
Request	17	573.47	437.936	9	1387
Total	54	699.00	745.597	9	3796

DMOC group	N	Mean (s)	Std. Deviation (s)	Min (s)	Max (s)
Coordination	30	658.0	438.81	36	1512
Site	8	1568.0	1437.44	62	3796
Base Hospital	8	340.0	238.54	78	703
Shelter	3	247.0	209.96	9	406
Logistics (L&R)	5	399.8	431.83	106	1138
Total	54	699.0	745.59	9	3796

Statistical test	*p*-value
ANOVA	0.020 *
Kruskal–Wallis test	0.356

**. The difference is significant at the 0.05 level*

Statistical test	*p*-value
ANOVA	0.030 *
Kruskal–Wallis test	0.026 *

**. The difference is significant at the 0.05 level*

Figure 3. Task processing time across task types and DMOC groups in the October exercise.

We estimated a censored regression model to examine the relationship between task processing time and multiple factors of group performance. In the October exercise, we measured the task processing time of 54 tasks as uncensored samples, and three unfinished unit task times were treated as right-censored. With regards to the independent dummy variables of the unit task type, we chose "Decision making" as the baseline for comparison. We chose the "Site" group as the baseline for the DMOC group dummy variables. Other independent variables were defined, as shown in Table 2. Robust standard errors were used in making inferences to address possible heteroscedasticity across observations. The result of censored regression analysis for the October exercise is summarized in Table 3. Coefficients of the three unit task types were negatively and statistically significantly different from that of the baseline of "Decision making". Thus, the task processing time spent on decision making took longer than the time spent on inquiry, report, and request. The "Base Hospital" and "Shelter" groups took less time to finish their tasks than the "Site" group. We considered possible impacts related to the mental stress of simultaneously processing multiple unit tasks. Given the limited human resources of each player group, as the number of unit tasks to be completed increases, average processing time per unit task may increase. Results did not show a statistically significant relationship between the number of simultaneously processed unit tasks by group, elapsed time, and their task processing time. However, the number of simultaneously processed unit tasks and the associated mental stress may affect the quality of responses, as we will discuss later. Since each DMOC exercise was conducted within one and a half hours, effects of working hours were not found in the statistical results. The positive coefficient of the variable "Stage 1" indicated that the tasks in Stage 1 were finished more quickly than those in Stage 2. Unit tasks in Stage 1 of the communication network were conducted by the first groups in the network who received injects from exercise controllers. Directly recognizing emergencies from injects may be one reason that the first groups finished tasks in less time than subsequent groups.

Table 3. Censored regression analysis for October exercise.

Variable	**Coefficient**	***p*-Value**
Constant	1746.0 **	0.000
DMOC groups		
Base Hospital Group	−1153.9 **	0.004
Shelter Group	−840.0 *	0.023
Logistics and Reinforcement (L&R) Group	−630.0	0.117
Coordination Group	−618.9	0.072
Communication domain (within DMOC)	482.5	0.053
Unit task type		
Inquiry	−1063.4 *	0.012
Report	−855.2 **	0.005
Request	−1186.3 **	0.008
Simultaneously processed unit tasks	35.0	0.326
Elapsed time	0.004	0.942
Stage (Stage 1)	−758.5 *	0.017
Number of observations	57	
Uncensored/right censored	54/3	
Wald chi-square (11)	29.39, $p = 0.002$	
Log pseudo likelihood	−425.221	

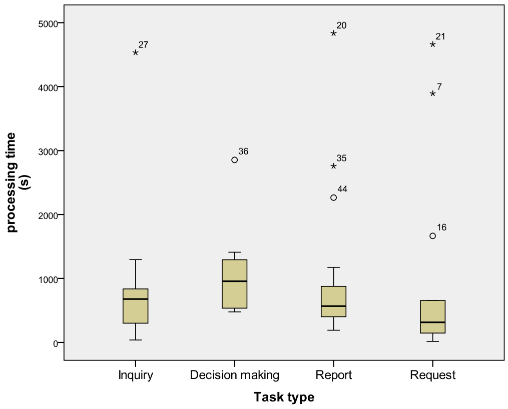
* $p \leq 0.05$; ** $p < 0.01$. Dummy-variable base line: Decision making and Site Group, Stage 2.

3.2.2. Analysis of the November Exercise

Figure 4 shows the results of applying the same analysis procedure to the November exercise. We measured the task processing time of 54 finished unit tasks and found five tasks unfinished until the end of the exercise. As we can see from the box plot, task processing time in the November exercise had more outliers than in the October exercise. Both ANOVA and Kruskal–Wallis tests did not detect statistically significant differences at the 5% level among the mean or median of task processing time across unit task types or DMOC groups. Censored regression analysis of the 59 tasks for this exercise did not identify any variables that were statistically significant at the 5% level. Thus, the time study of this research suggested that the characteristics of communications across the player groups changed between the October and November exercises. This difference between the two exercises was difficult to detect without an application of the time study. By comparing the mean processing time across task types between October and November exercises, "Decision making" in the November exercise consumed less time than in the October exercise. However, other task types in the November exercise used more time to complete than in the October exercise. One possible explanation is that groups make more efforts to conduct "Decision making" tasks, and then require longer processing time for completing other tasks. There were more outliners of time results in box plots of November data than October. Another possible reason is some new communication failures occurred during the November exercise. These explanations would be more convincing if we could investigate repeated tasks in

October and November exercises and interview players immediately after these exercises. In the next section, we attempt to identify some clues by comparing task processing time of 39 repeated unit tasks between the two exercises.

Task type

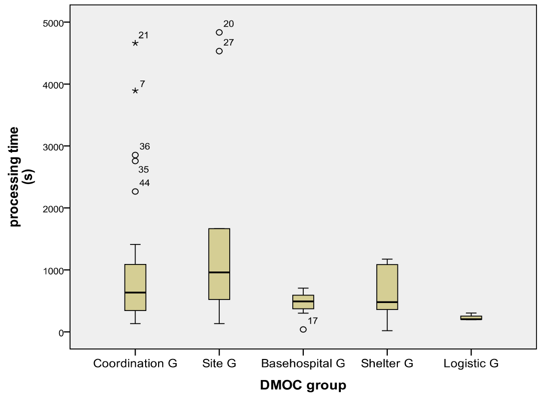


Types of task	N	Mean (s)	Std. Deviation (s)	Min (s)	Max (s)
Inquiry	10	983.70	1297.00	38	4532
Decision making	8	1113.25	782.40	479	2854
Report	22	929.68	1074.29	191	4834
Request	14	947.57	1475.06	17	4661
Total	54	971.52	1166.80	17	4834

Statistical test	*p*-value
ANOVA	0.986
Kruskal–Wallis test	0.097

*. *The difference is significant at the 0.05 level*

DMOC group



DMOC group	N	Mean (s)	Std. Deviation (s)	Min (s)	Max (s)
Coordination	28	1039.0	1169.52	132	4661
Site	10	1590.4	1687.29	133	4834
Base Hospital	8	456.3	207.26	38	704
Shelter	5	622.6	493.40	17	1174
Logistic (L&R)	3	234.0	59.02	196	302
Total	54	971.5	1166.80	17	4834

Statistical test	*p*-value
ANOVA	0.190
Kruskal–Wallis test	0.079

*. *The difference is significant at the 0.05 level*

Figure 4. Task processing time across task types and DMOC groups in November.

3.2.3. Analysis of Repeated Tasks in October and November Exercises

In Figure 5, we compared the difference in task processing time of 39 repeated unit tasks between the two exercises by subtracting the task processing time in November from that in October. Different markers indicate different DMOC groups. The figure shows 32 tasks in Stage 1 and seven tasks in Stage 2. Markers connected by a solid line indicate that these unit tasks were started by the same group at the same elapsed time.

Figure 5 reveals some curious results. In Stage 1, one task conducted by the “Coordination” group showed a significant difference in task processing time between the two exercises. This task was a response to inject 3, and the “Coordination” group was required to ask the “Logistics” group to send advance requests for reinforcement to the prefectural and national governments and to prepare an acceptance plan for these reinforcements. It took 856 and 3893 s to finish this task in October and November, respectively. In the November exercise, the “Coordination” group may have forgotten to perform this task for some time. This phenomenon may have occurred due to emergency states such as mental stress, due to heavy workloads among the “Coordination” group. In inject 3, the group had to conduct 11 unit tasks at the same elapsed time. The group were also in charge of conducting the highest number of tasks throughout the exercises.

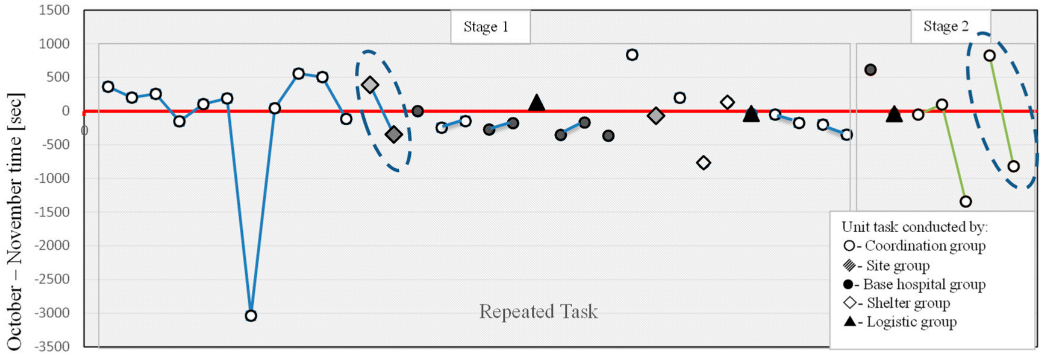


Figure 5. Differences in task processing time for 39 repeated unit tasks.

Figure 5 also reveals that the "Base" hospital group took more time to finish all tasks in Stage 1 in November than in October. One possible reason for longer time consumption in November is the group putting more efforts into providing detailed explanations during their communication with other groups. Evaluation tools in DMOC exercises were focused on timeliness of players to complete their unit tasks. From these specific findings, conducting interviews with player groups may highlight root causes for this phenomenon. It helps to comprehensively evaluate the performance of participants.

There are cases in which the processing time of unit tasks started at the same time by the same group revealed reversed results regarding time difference between October and November (see circled results in Figure 5). The last two unit tasks on the right-hand side of Figure 5 are an example of this. A detailed description of this unit task in response to inject code 19 in October is shown in Figure 1. In October, Unit task 2.1 "Contact headquarters team in EOC" and Unit task 2.2. "Report status to the firefighting dispatch center" were finished in 1442 and 1449 s, respectively. In November, Unit task 2.1 took 615 s, and Unit task 2.2 took 2265 s. The total task processing time for these two unit tasks was approximately equal in the October and November exercises. This reverse of processing time difference between the two months may be a result of a change in prioritization of these unit tasks by the players. Thus, we were able to reveal clues about players' thoughts and behavior by conducting this detailed time analysis of exercise results.

4. Conclusions and Further Implications

4.1. Conclusions

Functional exercises are an important type of training that enhance information sharing through networks among emergency response organizations. Evaluation is an essential part of exercise management programs to consider how an exercise met its objectives and how emergency work was performed by players. Other benefits in exercise programs include promoting common understanding of the current status of the emergency management organization and providing lessons learned to encourage organizational improvement. In practice, checklists, questionnaires, and interviews are common tools for collecting information for exercise results. Our study gave greater consideration to processing time of unit tasks to develop methods for quantitatively evaluating group performance in public health service functional exercises. In a functional exercise, these evaluation methods can be used separately or together with conventional methods, depending on exercise objectives. In DMOC exercises, one of the main objectives for medical headquarters' teams in Kitakyushu was to improve efficiency of communication among groups. Therefore, it was significant to measure the value of task processing time of DMOC groups.

We used knowledge of time studies in the industrial domain, communication studies, and team performance studies to interpret the meaning of task processing time observed in DMOC exercises. Our time analysis found several features of players' communications that would have been difficult to identify without the method. The task processing time in "Decision making" and "Site" groups was longer than other unit task types and in other groups in October. Unit tasks implemented in Stage 1 consumed less time than tasks in Stage 2. These differences, however, disappeared in November with more outliers in box plots. We provided some discussions and possible reasons for this phenomenon. We compared processing times of the same unit tasks between the October and November exercises, which revealed communication failure and change in prioritization over tasks between two exercises.

This study provides a new evaluation method for functional exercises using time data. Examining the relationship between task processing time and multiple factors of group performance in an exercise reveals more evidence for better understanding of the exercise. Identification of some specific issues on participants' performances during exercises proved to be difficult without this method. Sharing these issues in the hotwash section of an exercise would be beneficial for discussion among participants. Results from these activities provide useful information for emergency managers to understand the current status of his/her organization and improve emergency planning.

Another benefit of the time-based evaluation method used in this study is that it enables longitudinal data for comparing and analyzing repeated exercises, which contributes to continuous improvement of exercise programs. In the standard condition, the length of time for processing each unit task in an exercise may be used to determine a standard unit processing time for completing such unit tasks and a deadline for the entire workflow. This information on unit task processing time would help make standard operating procedures of emergency management more practical.

4.2. Limitations and Further Study

There are some limitations that should be improved in future research. Our analysis method would be more robust if the exercises had a larger number of injects. In our case study of DMOC exercises, the characteristics of group members were not observed. Information on the skills and characteristics of members in each group should be obtained before the exercise to enable better comparison among groups. If we could have a meeting with the players of the exercise to identify root causes of the time-related issues found through our analysis, we could have more explanations to justify the phenomenon in DMOC exercises. Providing an opportunity to discuss time analysis results with exercise participants and identifying its effects on organizational improvement is necessary for future exercises. The simulated emergency in the DMOC exercises was different from real large-scale emergencies in some aspects. Exercise time was short and workloads were not heavy for some groups. Exercise players were not subjected to real risks and there was little chance of panic to happen. By changing the setup of functional exercises, we can obtain additional information on the effects of stressors on unit task processing times. For example, we will be able to see how fatigue affects unit task processing times in a functional exercise that lasts for several days.

Since an exercise evaluation program is designed to achieve the objectives of the training, the best evaluation methods should be selected in accordance with the training objectives. We used data of DMOC exercises that were designed to focus on timeliness of players to complete their unit tasks. In an emergency response, how well tasks are performed, and the speed of execution are both important. Time is a clear metric to describe outcomes, but it does not comprehensively describe the quality of an action. However, without time measurement, large data from exercise programs would not be revealed for organizational improvements. Therefore, by considering the time element together with other information obtained from checklists, the evaluator's comments, and open-ended questions, emergency managers can be more accurate in answering the following four questions: what happened, why and when it happened, and how it should be improved through exercise programs.

Author Contributions: Conceptualization, N.T.T., T.K., M.T., and K.K.; methodology, N.T.T.; Validation, T.K., M.T. and K.K.; formal analysis, N.T.T. and M.I.; investigation, N.T.T. and M.I.; resources, S.I.; writing—original draft preparation, N.T.T.; writing—review and editing, T.K.; supervision, S.I.; project administration, S.I.; funding acquisition, T.K.

Funding: This research was funded by the Japan Society for the Promotion of Science, KAKENHI Grant Number 16K12840.

Acknowledgments: The authors appreciate DMOC members for providing the opportunity to measure unit task-time in their exercises. The study authors obtained Ethic Approval from the University of Kitakyushu, Japan, Ethics Committee Regarding Research on Human Beings Application Number 19-5.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Haddow, G.D.; Bullock, J.A.; Coppola, D.P. *Introduction to Emergency Management*, 5th ed.; Elsevier: Waltham, MA, USA, 2014; pp. 1–4, ISBN 978-0-12-407784-3.
2. Lettieri, E.; Masella, C.; Radaelli, G. Disaster management: Findings from a systematic review. *Disaster Prev. Manag. Int. J.* **2009**, *18*, 117–136. [CrossRef]
3. Coppola, D.P. *Introduction to International Disaster Management*; Elsevier: Waltham, MA, USA, 2006; pp. 1–15, ISBN 0-08-046573-0.
4. Alexander, D. *Principles of Emergency Planning and Management*; Oxford University Press: Oxford, NY, USA, 2002; pp. 190–215, ISBN 978-0-19-521838-1.
5. McLay, L.A.; Mayorga, M.E. Evaluating emergency medical service performance measures. *Health Care Manag. Sci.* **2010**, *13*, 124–136. [CrossRef] [PubMed]
6. Kobusingye, O.C.; Hyder, A.A.; Bishai, D.; Joshipura, M.; Hicks, E.R.; Mock, C. Emergency medical services. *Dis. Control Priorities Dev. Ctries.* **2006**, *2*, 626–628.
7. Dausey, D.J.; Buehler, J.W.; Lurie, N. Designing and conducting tabletop exercises to assess public health preparedness for manmade and naturally occurring biological threats. *BMC Public Health* **2007**, *7*, 92. [CrossRef] [PubMed]
8. Skryabina, E.; Reedy, G.; Amlot, R.; Jaye, P.; Riley, P. What is the value of health emergency preparedness exercises? A scoping review study. *Int. J. Disaster Risk Reduct.* **2017**, *21*, 274–283. [CrossRef]
9. Nazli, N.N.N.N.; Sipon, S.; Radzi, H.M. Analysis of training needs in disaster preparedness. *Procedia Soc. Behav. Sci.* **2014**, *140*, 576–580. [CrossRef]
10. Klima, D.A.; Seiler, S.H.; Peterson, J.B.; Christmas, A.B.; Green, J.M.; Fleming, G.; Thomason, M.H.; Sing, R.F. Full-scale regional exercises: Closing the gaps in disaster preparedness. *J. Trauma Acute Care Surg.* **2012**, *73*, 592–597. [CrossRef] [PubMed]
11. National Incident Management System. Available online: https://www.fema.gov/national-incident-management-system (accessed on 25 June 2019).
12. International Atomic Energy Agency (IAEA). *Preparation, Conduct and Evaluation of Exercises to Test Preparedness for a Nuclear or Radiological Emergency*; International Atomic Energy Agency: Vienna, Austria, 2005.
13. Quarantelli, E.L. Disaster crisis management: A summary of research findings. *J. Manag. Stud.* **1988**, *25*, 373–385. [CrossRef]
14. Russo, F.; Rindone, C. Civil risk manager at European level. *WIT Trans. Ecol. Environ.* **2013**, *173*, 765–778.
15. Muttarak, R.; Pothisiri, W. The Role of Education on Disaster Preparedness: Case Study of 2012 Indian Ocean Earthquakes on Thailand's Andaman Coast. *Ecol. Soc.* **2013**, *18*, 51. [CrossRef]
16. Homeland Security Exercise and Evaluation Program (HSEEP). Available online: https://www.fema.gov/hseep (accessed on 8 June 2019).
17. Annelli, J.F. The national incident management system: A multi-agency approach to emergency response in the United States of America. *Rev. Sci. Tech. Int. Off. Epizoot.* **2006**, *25*, 223–231. [CrossRef]
18. Militello, L.G.; Patterson, E.S.; Bowman, L.; Wears, R. Information flow during crisis management: Challenges to coordination in the emergency operations center. *Cogn. Technol. Work* **2007**, *9*, 25–31. [CrossRef]
19. Mendonça, D.; Fiedrich, F. Training for improvisation in emergency management: Opportunities and limits for information technology. *Int. J. Emerg. Manag.* **2006**, *3*, 348–363. [CrossRef]
20. Phelps, R. *Emergency Management Exercises: From Response to Recovery: Everything You Need to Know to Design a Great Exercise*; Chandi Media: San Francisco, CA, USA, 2011; pp. 10–18, ISBN 0-9831143-0-7.
21. Johnson, Y.J.; Herrmann, J.A.; Wallace, R.L.; Troutt, H.F.; Myint, M.S. Development and Implementation of a Functional Exercise to Assess Public Health Agency Response to Foodborne Bioterrorism. *J. Homel. Secur. Emerg. Manag.* **2009**, *6*, 49. [CrossRef]

22. Savoia, E.; Agboola, F.; Biddinger, P. A Conceptual Framework to Measure Systems' Performance during Emergency Preparedness Exercises. *Int. J. Environ. Res. Public. Health* **2014**, *11*, 9712–9722. [CrossRef] [PubMed]
23. Gebbie, K.M.; Valas, J.; Merrill, J.; Morse, S. Role of exercises and drills in the evaluation of public health in emergency response. *Prehospital Disaster Med.* **2006**, *21*, 173–182. [CrossRef] [PubMed]
24. Russo, F.; Rindone, C. Urban exposure: Training activities and risk reduction. *WIT Trans. Ecol. Environ.* **2014**, *191*, 991–1001.
25. Canton, L.G. *Emergency Management: Concepts and Strategies for Effective Programs*; Wiley-Interscience: Hoboken, NJ, USA, 2007; pp. 116–124, ISBN 978-0-471-73487-1.
26. Harvard School of Public Health. *Public Health Emergency Preparedness Exercise Evaluation Toolkit*; Harvard School of Public Health: Boston, MA, USA, 2013.
27. Nelson, C.; Lurie, N.; Wasserman, J. Assessing public health emergency preparedness: Concepts, tools, and challenges. *Annu. Rev. Public Health* **2007**, *28*, 1–18. [CrossRef] [PubMed]
28. Vukmir, R.B. Survival from prehospital cardiac arrest is critically dependent upon response time. *Resuscitation* **2006**, *69*, 229–234. [CrossRef] [PubMed]
29. Sanchez-Mangas, R.; Garcia-Ferrrer, A.; De Juan, A.; Arroyo, A.M. The probability of death in road traffic accidents. How important is a quick medical response? *Accid. Anal. Prev.* **2010**, *42*, 1048–1056. [CrossRef] [PubMed]
30. California Department of Fish and Game. *Drills & Exercises and Evaluation Guidance Manual*; California Department of Fish and Game: Sacramento, CA, USA, 2010.
31. Center for Health Policy. *Public Health Emergency Exercise Toolkit*; Columbia University: Columbia, NY, USA, 2006.
32. WebEOC. EPA OSC Response. Available online: https://response.epa.gov/_help/WebEOCUserGuide.pdf (accessed on 25 April 2019).
33. Kohriyama, K.; Shigehito, I.; Takaaki, K. Saigai iryo ni hitsuyona hinichijyosei (Awareness of unusual situations in disaster medical management). *Risk Taisaku Dot Com.* **2016**, *53*, 82–84. (In Japanese)
34. Jain, S. A parametric approach to emergency response project management. In Proceedings of the PMI Research Conference: New Directions in Project Management, Montreal, QC, Canada, 19 July 2006.
35. Fiedrich, F.; Gehbauer, F.; Rickers, U. Optimized resource allocation for emergency response after earthquake disasters. *Saf. Sci.* **2000**, *35*, 41–57. [CrossRef]
36. Simonovic, S.P.; Ahmad, S. Computer-based Model for Flood Evacuation Emergency Planning. *Nat. Hazards* **2005**, *34*, 25–51. [CrossRef]
37. Kitakyushu Fire Department. *Kitakyushu Functional Exercise After Action Report*; Kitakyushu Fire Department: Kitakyushu, Japan, 2008. (In Japanese)
38. Taylor, F.W. *The Principles of Scientific Management*; Harper & Brothers: New York, NY, USA, 1911; pp. 1–77.
39. Payne, S.C.; Youngcourt, S.S.; Watrous, K.M. Portrayals of F.W. Taylor across textbooks. *J. Manag. Hist.* **2006**, *12*, 385–407. [CrossRef]
40. Monden, Y. *Toyota Production System: An Integrated Approach to Just-in-Time*, 4th ed.; CRC Press: Boca Raton, FL, USA, 2012; pp. 137–177. ISBN 978-1-4398-2097-1.
41. Annett, J.; Duncan, K.D. Task analysis and training design. *J. Occup. Psychol.* **1967**, *41*, 211–221.
42. Annett, J.; Stanton, N.A. *Task Analysis*; CRC Press: Boca Raton, FL, USA, 2000; pp. 9–24, ISBN 978-1-4200-2298-8.
43. Habermas, J. What is universal pragmatics? In *Communication and the Evolution of Society*; McCarthy, T., Translator; Beacon Press: Boston, MA, USA, 1979; pp. 1–68.
44. McGrath, J.E. *Social Psychology: A Brief Introduction*; Holt, Rinehart and Winston: London, UK, 1970; pp. 10–28. ISBN 978-0-03-910078-0.
45. Gladstein, D.L. Groups in Context: A Model of Task Group Effectiveness. *Adm. Sci. Q.* **1984**, *29*, 499. [CrossRef]
46. Hackman, J.R. The design of work teams. In *Handbook of Organizational Behavior*; Lorsch, J., Ed.; Prentice-Hall: Englewood Cliffs, NJ, USA, 1987; pp. 315–342.
47. Jex, S.M.; Britt, T.W. *Organizational Psychology: A Scientist-Practitioner Approach*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2014; pp. 411–436, ISBN 978-1-118-72407-1.
48. Ministry of Health Malaysia. *Manual on Mental Health and Psychosocial Response to Disaster in Community*; Ministry of Health Malaysia: Putrajaya, Malaysia, 2013.

49. Kowalski-Trakofler, K.M.; Vaught, C.; Scharf, T. Judgment and decision making under stress: An overview for emergency managers. *Int. J. Emerg. Manag.* **2003**, *1*, 278. [CrossRef]
50. Hartsough, D.M.; Myers, D.G. *Disaster Work and Mental Health: Prevention and Control of Stress among Workers*; National Institute of Mental Health: Rockville, MD, USA, 1985.
51. Mitchell, J.T.; Dyregrov, A. Traumatic stress in disaster workers and emergency personnel. In *International Handbook of Traumatic Stress Syndromes*; Springer: Berlin/Heidelberg, Germany, 1993; pp. 905–914.
52. Janis, I.L. Decision making under stress. *Handb. Stress Theor. Clin. Asp.* **1982**, *2*, 56–74.
53. Driskell, J.E.; Salas, E. Group decision making under stress. *J. Appl. Psychol.* **1991**, *76*, 473. [CrossRef]
54. Driskell, J.E.; Salas, E.; Johnston, J. Does stress lead to a loss of team perspective? *Group Dyn. Theory Res. Pract.* **1999**, *3*, 291. [CrossRef]
55. Poulton, E.C. Arousing environmental stresses can improve performance, whatever people say. *Aviat. Space Environ. Med.* **1976**, *5*, 1193–1204.
56. West, R.; MacDougal, K.; Ward, L. Analyzing fatigue, stress and human errors in Emergency Operation Centre management: The consequences of using different cognitive modelling frameworks. In Proceedings of the 14th International Conference on Cognitive Modeling, University Park, PA, USA, 13 July 2016.
57. Brown, C.; Hearty, L.; Heideman, M.; Moran, C.; Pearce, L. *Guidebook for Team Support Workers in Emergency Operation Centers*; Justice Institute of British Columbia: New Westminster, BC, Canada, 2015.
58. Maddala, G.S. *Limited-Dependent and Qualitative Variables in Econometrics*; Cambridge University Press: Cambridge, UK, 1983; pp. 149–182, ISBN 978-0-511-81017-6.