Journal of Homeland Security and Emergency Management

Volume 5, *Issue* 1 | 2008 | *Article* 37

Longitudinal Evaluation of Food Safety Discussion-Based Exercises: Tool Development and Initial Validation

Elizabeth A. Prebles*
Aylin D. Sayir†
Dale C. Brandenburg‡
Edward C. Mather**

*Wayne State University, ax1695@wayne.edu
†Karmanos Cancer Institute, aylin.sayir@gmail.com
‡Wayne State University, d.brandenburg@wayne.edu
**Michigan State University, mather@cvm.msu.edu

Longitudinal Evaluation of Food Safety Discussion-Based Exercises: Tool Development and Initial Validation*

Elizabeth A. Prebles, Aylin D. Sayir, Dale C. Brandenburg, and Edward C. Mather

Abstract

This study was undertaken for three major purposes: first, to provide initial validation data on an assessment tool designed to provide evidence supporting the following two purposes; second, to analyze the lessons learned process from exercise conduct through subsequent organizational action; and third, to investigate the long-term impact of discussion-based exercises associated with improving food safety preparedness. The major hypothesis forming the basis of the investigation was a logic model containing five sets of variables selected from relevant literature on emergency/disaster preparedness, training evaluation, and instructional design. One variable set, Organizational Actions, served as the criterion of interest. The other four variables served as predictor variables; these were: a) Exercises Outcomes; b) Exercise Design Features; c) Organizational and Personal Characteristics, and d) Demographic Characteristics. The target exercises selected for investigation and evaluation were a class of discussion-based activities conducted for various audiences of food safety personnel from government, industry, public health, and academia. The four exercise sessions from which participants were selected had occurred between two years and six months previously. There were two stages of data collection—personal interviews to assure data relevance and an online survey. Fourteen participants were interviewed and 47 complete responses were obtained from the survey. Data analysis provided considerable support for assessment tool validity. First, it would appear that discussion-based exercises can have some positive long-term impact on organizational change, a finding that heretofore, has not been demonstrated in the relevant research literature. Second, empirical evidence was found for the logic model, with almost

*This research was supported by the U.S. Department of Homeland Security (Grant number N-00014-04-1-0659), through a grant awarded to the National Center for Food Protection and Defense at the University of Minnesota and a sub-grant (R9106000101) awarded to Michigan State University and Wayne State University. Additional support for Dr. Brandenburg was supplied by the National Science Foundation (NSF Grant # 0428216) to Wayne State University. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of Homeland Security or the National Science Foundation. The cooperation and support of the Uriah Group, the designers of the exercises evaluated in this report, is gratefully acknowledged. Requests for further information should be directed to Dr. Brandenburg.

60% of the variance in the criterion variable accounted for by the other measures. Third, given that the exercises were designed to promote coordination and collaboration across agencies and industry sectors, it was found that by tailoring the exercise to participant backgrounds and designing activities to promote long-term learning, the potential for cross-organizational action can be increased. Other conclusions highlighted specific recommendations that could improve exercise design.

KEYWORDS: exercise evaluation, emergency response coordination, lessons learned, organizational impact

INTRODUCTION

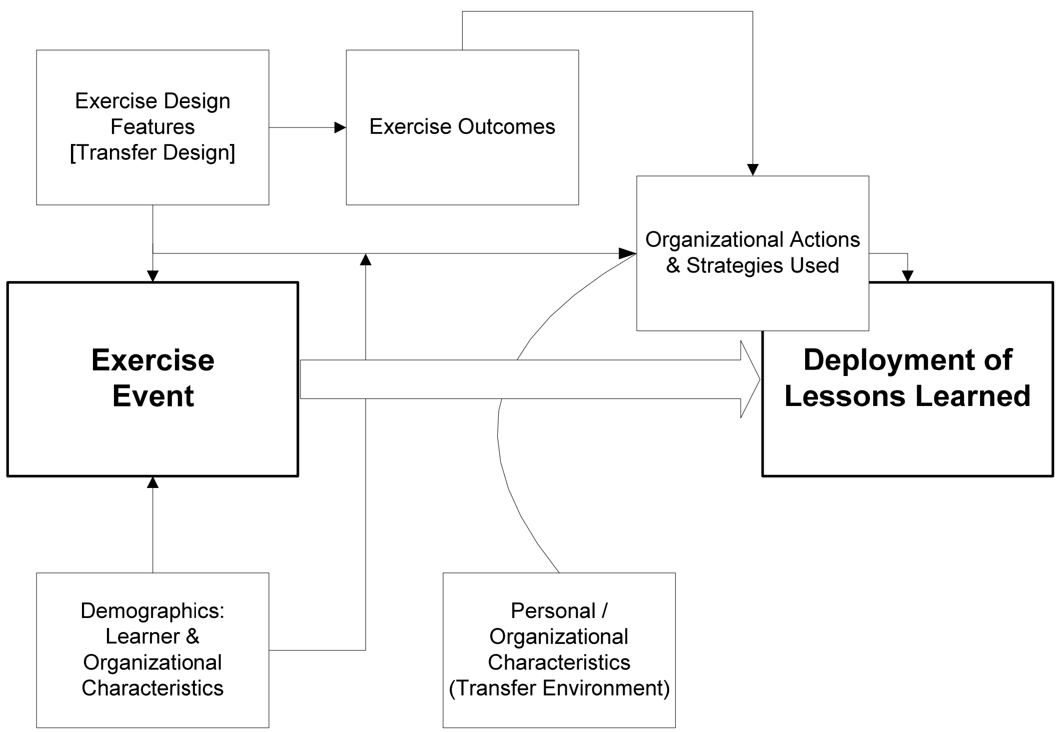
Within the field of emergency and disaster preparedness, the general intent of discussion-based exercises is often limited to awareness of concepts, terminology, procedures, and resources. However, when such exercises are designed to increase collaboration and cooperation among multiple agencies or industry sectors, the outcomes of such exercises can have important long-term impacts. A standard approach in most emergency and disaster preparedness exercises is to collect "lessons learned", typically important conclusions from the exercise that can then be deployed within the participating organizations. These lessons learned can assist organizational change efforts which in turn improve the response preparedness capability of an organization.

While this approach appears logical, there is limited empirical evidence to suggest how "lessons learned" get translated into future organizational actions. This study represents one effort to empirically 'connect the dots' from the exercise event to organizational action. Such action cannot be measured immediately after the exercise or training event, but after some time period has elapsed. To examine the lessons learned process requires a longitudinal study of selected key variables hypothesized to influence this process. These variables form the basis of an assessment tool designed to measure the long-term effectiveness of discussion-based exercises and are based on the following requirements:

- The data collected will test a hypothesized logic model depicting how five selected variable sets interact to provide empirical evidence to assist in understanding the lessons learned process;
- Evaluation issues will be focused upon issues of coordination across individuals and organizations participating in a discussion-based exercise;
- The tool will measure longitudinal impact or organizational actions attributed to exercise participation especially to assess the value for participating stakeholders;
- The study outcomes will contribute to improving the process of learning from this training so that future exercises are more effective and efficient.

The target training event for this investigation was a special type of discussion-based exercise designed to provide information and concepts related to protection of food safety, and it contains a number of elements focused on inter-agency coordination. This design is called a Reality-Based Exercise (RBX™), and it is a product of the Uriah Group. The RBX™, in differing from the traditional discussion-based format, is designed to simulate real world chaos in a structured format to promote discussion among the key stakeholders (typically separated by disciplines) regarding a hypothetical incident, stress the management and creation of policies and procedures for such an occurrence, and from a post-exercise perspective, encourage changes in the participant's organization. A typical exercise run takes about six weeks of preparation that includes the formation of a steering committee of local stakeholders to guide the process, shooting of local scenes to provide video injects associated with a simulated crisis event, discipline-specific situation manuals, exercise evaluation guidelines, and two versions of an After Action Report. The exercise session is conducted in six hours, using up to 80 injects (more than three times the number typically used), and forces discipline-specific decisions which must then be reconciled with decisions among other discipline groups.

Figure 1
Logic Model for Longitudinal Effectiveness of Exercises



The five components of the evaluation process are summarized in the Logic Model diagram of Figure 1. A logic model, in general, is meant to portray the major assumptions and hypotheses that associate the evaluation with the underlying research issues. This logic model was also meant to depict the process of organizational learning, that is, the theoretical constructs that are designed to be captured through the data collection. In this case, we make the general assumption that the learning event, here the Reality-Based Exercise™ session, has a direct influence on the ultimate outcome, the Deployment of Lessons Learned (as shown by the large arrow connecting the two end points of the model). The major criterion variable, Deployment of Lessons Learned, was not measured directly because this would entail field-based performance data captured as a result of an actual emergency or disaster. Instead, we will use Organizational Actions as an intermediate outcome and surrogate for this measure because it is assumed to have substantive components in common with the major criterion. The other variable sets are assumed to influence Organizational Actions and the extent of those influences is tested in this study.

The following discussion outlines the research variables underlying the development of the assessment tool depicted in the logic model.

1. *Organizational Actions.* If the exercise is effective over the long term, the organization needs to implement some actions or strategies (Drabek, 2005; Drabek & McEntire, 2002) that complemented the objectives. Such actions can be derived from the use of an After Action Report, notes taken by participants, ideas generated during the exercise discussion, or a list of "lessons learned" generated at the conclusion of the exercise.
2. *Exercise Outcomes (immediate)*. The vast majority of discussion-based exercises list more learning objectives than can be typically accomplished within the three to six-hour period allocated for the event. As a result, only a few objectives can be effectively accomplished. Thus, for this exercise format, we are interested in determining which short-term or immediate objectives were accomplished across the selected settings. Since the majority of the objectives were associated with cross-organizational coordination, a set of objectives associated with coordination outcomes (Drabek, 2005) were selected from the research and practitioner literature.
3. *Organizational / Personal Characteristics (Transfer Environment).* Opportunity to change "lessons learned" into organizational actions are hypothetically linked to certain characteristics of the learner and the organization (Holton, 2005). In other words, organizational context or culture, especially the culture of organizational learning, is hypothesized to influence whether or not the lessons are actually "learned". Personal incentives, supervisor support, and resource allocation are some of the factors that contribute to this environment.
4. *Exercise Design Features.* For an exercise, or any learning event, to be effective in the long term, the design of the exercise should have some particular characteristics that promote this long term learning (Holton, 2005). We are distinguishing the immediate reaction to an exercise found from the 'smile sheets' typically collected at the conclusion of the exercise from those aspects that could have a lingering effect on subsequent learning. Features such as problem-solving strategies, relevance of the

scenario to job environment, and types of decisions made are representative learning design elements that can be built into the exercise to assist in transferring immediate learning into longer term impact.

5. *Demographic Characteristics*. It can be expected that individuals who have leadership positions in organizations, or have influence in their work groups, or who have recognized experience in the field have a greater likelihood of influencing the lessons learned process (Gillespie & Streeter, 1987). It is also possible that the type of organization, public or private, may also influence the effective implementation of learning.

Research Questions

Taken together, this study aims to uncover several questions regarding the logic model and its components.

1. Does the logic model suggest the ability to predict organizational action?
2. Are the organizational actions related to the other four subscales (exercise outcomes, organizational/personal characteristics, exercise design features, demographic characteristics)?
3. Is there practical significance associated with identifying key item relationships with the organizational action criterion?

REVIEW OF RELATED LITERATURE

Currently there exist few published investigations related to the longitudinal evaluation or impact of discussion-based exercises. Thus, one aim of this project was to add to the literature and promote further studies on this important topic. Concomitantly, the lessons learned process associated with emergency and disaster preparation exercises is not well understood and this project seeks to further knowledge regarding this process. Based upon the disaster and emergency response literature, this study sought also to understand and improve discussion-based exercises as well as provide evidence to better understand the impact exercises on individuals and organizations, especially when associated with coordination of preparation and response. Such coordination is critical in the food safety domain especially given the diversity of the supply chain (represented as 'farm to fork'). With the advent of the bioterrorism initiatives of the Centers for Disease Control and Prevention beginning in 2002, and industry assessment programs from the US Department of Agriculture (2005), many activities have sought the cooperation of industry and government agencies. The working relationship for private and public organizations to develop a comprehensive system of food safety is still in its formative stages, especially with regard to intentional threats to the food supply chain (Kinsey, 2007).

Interest in coordination for public and governmental emergency and disaster response organizations is extensive. Since September, 11, 2001 this interest has spread to the preparedness and response for businesses. Preparedness has been defined in a variety of ways. In a review of conceptualizing preparedness, Gillespie and Banerjee (1993) note that preparedness activities can range from monitoring to warning systems to planning. Further, preparedness can be directly assessed if pre-disaster measures of preparedness tasks have been measured, or, indirectly assessed through disaster-relevant training and exercises (Gillespie & Banerjee, 1993). Webb, Tierney, and Dahlhamer (2000) suggest that as a result of increased coordination with local governmental preparedness organizations, businesses will incur a less severe impact and will show the most improvement following a disaster.

Webb, Tierney, and Dahlhamer (2000) acquired data from 5000 private sector firms in communities in the United States regarding hazard awareness, preparedness, disaster impacts, and recovery. They cite several factors influencing business preparedness and the adoption of other self-protective measures. The researchers found that in the Memphis/Shelby county businesses and the Northridge earthquakes, business disaster preparedness was quite low. This data is similar to research conducted by Drabek (1994), suggesting the 180 tourist firms he studied had unsatisfactory business disaster preparedness efforts, and Mileti, Darlington, Fitzpatrick, and O'Brien (1993), highlighting low levels of preparedness in the San Francisco Bay area businesses. Webb, Tierney, and Dahlhamer (2000) also found that of the business preparedness efforts, retail and service sectors, such as restaurants and grocery stores, tend to lag.

How then do organizations realize the value of preparedness and coordination? Drabek and McEntire (2002) suggest first examining emergent response processes. They state inter-organizational networks must first agree that something needs to be done in order to increase coordination. Emergence is "the process through which organizational personnel dedicate and structure themselves as they seek to resolve the demands placed on their community in times of disaster" (Drabek & McEntire, 2002, p. 198). The process of emergence then leads to multi-organizational collaboration. This collaboration implicates a need for effective coordination strategies (Drabek, 2001).

Drabek and McEntire (2002) define coordination as "collaborative processes through which multiple organizations interact to achieve common objectives," (p. 199). Coordination is based on pre-disaster understanding facilitated by planning and training (Drabek & McEntire, 2002). Another definition offered by Quarantelli states coordination is "mutually agreed upon cooperation about how to carry out particular tasks" (1997, p. 48). Important to note, however, is that coordination efforts do not guarantee coordination will result, and there are varying degrees of coordination (Drabek & McEntire, 2002).

Auf der Heide (1989) reveals that a lack of coordination may be attributed to, among other things, factors such as misunderstandings among organizations, failure to fully utilize equipment and personnel, omission of essential tasks, and overly taxed procedures. The very nature of disasters themselves makes them difficult to prepare for and coordinate within. For example, Auf der Heide (1989) states that disasters are difficult to prepare for due partially to the existence of cross-jurisdictional boundaries, the need to undertake unfamiliar tasks, the potential to trigger mobilization of participants that do not ordinarily respond, and the inherent change in structures of responding organizations. In addition to Auf der Heide's argument, several scholars cite the decentralized nature of the US political system, intergovernmental and intra-organizational complexities make disaster preparedness and coordination difficult (Petak, 1985; Toulmin et al., 1980; Waugh, 1993; Drabek, 1985; Quarantelli, 1997). The difficulty coordination poses in disaster preparedness calls for an improvement in the overall strategies required to implement coordination.

Drabek and McEntire (2002) created an extensive review of literature regarding improving coordination and preparedness. Ultimately, they cite scholars who suggest organizations working together before a disaster withstand fewer problems during a disaster (Auf der Heide, 1989; Tierney, 1994; Carley & Harrald, 1997; Drabek, 2005). Specifically, Drabek (1985; 2005), Auf der Heide (1989), and Quarantelli (1997) suggest training and exercise to increase disaster preparedness and coordination. The benefit of training and exercise is evident in creating contacts and increased communication between organizations (Wegner, 1989; Quarantelli, 1997; Drabek, 2005). Paton et al. (1998) additionally suggest promoting coordination vis-à-vis pre-disaster networking. The relevance of these pre-disaster measures cannot be realized without an evaluation measure.

The importance of evaluation is undisputed (Potter, Ley, Fertman, Eggleston, & Durman, 2003). The authors found that a "thorough and comprehensive evaluation plan can ensure that all constituents receive needed data and information (p. 493)." Despite this, most training programs, even large-scale ones such as the RBX™, are not evaluated to determine their long-term effectiveness. Holton (2005) suggests a human resource developmental model for evaluation and research as a comprehensive framework diagnosing and understanding the causal influences of existing evaluation models. His model suggests several points of consideration pertinent to this study; namely, the value and use of the lessons learned process within organizations. Holton suggests that perceptual measures of training effectiveness may have a role in predicting learning and performance outcomes. One way to measure the perceptual measures of training is to analyze the lessons learned data. According to Holton (2005), two primary factors that influence the lessons learned process are transfer design and transfer context. Transfer design is associated with how the training or exercise plan is developed. Such design elements would include learning objectives, expected outcomes, and the strategies used to instruct the learners. Use of problem-solving in a chaotic environment, as purported within the RBX™ exercise, would be an example of transfer design whereby learners can “transfer” the result to a future real event. Transfer context, reflects the organization’s capacity to accept the lesson and change in a way to better respond in the future. Obtaining management support for change, getting the right staff into the training, and having a culture where new strategies are accepted are examples of an environment (transfer context) where lessons learned can be implemented.

METHOD AND RESULTS

To determine the longitudinal effectiveness of RBX™s, four separate food safety exercise sessions conducted by the Uriah Group were selected for study. These exercises were chosen because they were conducted at varying points in time -- two years ago, a year and a half ago, a year ago, and six months ago from the onset of our data collection. Each was conducted for groups in different locations across the country. This project utilized a mixed methods design using both qualitative and quantitative data collection methods. There were two phases of this project – telephone interview followed by an online survey.

Method and Results of Telephone Interviews

The telephone interviews were conducted with randomly selected participants from each of the four sessions. The interview acted as a pilot study to provide information that would assist with the development of the online survey. Selected participants received an email explaining the project. One interview protocol was created and used by three interviewers. Each interviewer was trained on this protocol to ensure professionalism and consistency across all interviews. Each interview lasted approximately 20 minutes. The protocol was designed to elicit general knowledge and opinion of the discussion-based exercise and its effectiveness for personal and organizational development.

Fourteen interviews were collected and then subjected to a constant comparison method, a variant of grounded theory. Three constant comparative methods were performed: a comparison by interviewee, a comparison by question, and a comparison by region of exercise. The final constant comparison was used primarily to assess the impact of hedging, but it did not produce any new themes and reached a point of redundancy.

From the results of the telephone interviews, two significant encouraging results can be noted. One result indicated that the major collaborative objectives of the exercise, associated with increasing inter-organizational interaction, were fulfilled and that follow-on actions demonstrated the value for increased response capacity. A second result showed that organizations recognized the need to understand and fully document internal capabilities in order to improve response effectiveness. Another theme that emerged from the interview data is the need for the After Action Report (AAR) to be shared with exercise attendees. Those respondents whose organization's shared the After Action Report and reviewed it as a post-exercise activity expressed a great appreciation its availability. It appears from this data that the AAR helps organizations to have a lessons learned session after the exercise. This serves as a time of follow-up and planning.

One lesson learned from the AAR was that it highlighted an organization's communication gaps. Participants who had access to and reviewed the AAR were better able to recall the lessons from the exercise and what their organization did to carry out these lessons. Unfortunately, a majority of participants reported that they did not receive or gain access to an AAR.

Overall, participants expressed that the creation, maintenance, and use of contact lists was the most relevant piece of the exercise. Another common theme expressed by participants was the use of training and communication to facilitate preparation and response as well as increasing organizational collaboration.

Methods and Results of Online Survey

This survey was administered to randomly selected participants from each of the four sessions. Each selected participant received an email explaining the project and containing a link to the survey. A total of 70 out of 230 selected respondents provided some data whereas 47 respondents provided complete data on most items. Thus, the general response rate was 30%, which is considered a high rate of response in the emergency and disaster research field, but the completion rate was 20% based on the entire sample. To retain as many valid responses as possible, a respondent who did not reply to a particular item was considered as missing data for only that item as opposed to eliminating the respondent from all analyses. Due to the limited sample size, the demographic variables were recoded into a smaller number of categories. One of four groups sampled had a small number of returns and another also had less than 10 responses eliminating the possibility of comparing results across groups. Government represented the largest participant organization type with 18; industry and the 'other' category had the same number of participants with 12 each. The 'other' category was inclusive of academia and public health as well as those unlisted in this survey. Almost half of the respondents had more than 10 years of experience in the field of food safety while less then ten respondents had less then three years. Almost three-fourths of the respondents were managers with the remainder in staff positions. About twothirds did not have access to an after action report following participation in the training session.

Four sets of variables, or subscales were used to collect the major portion of the respondent data. Their summary labels are: Organizational Actions, Exercise Outcomes, Exercise Design Features and Organizational/Personal Characteristics. The highest ranked items in each of the subscales are displayed in Table 1 along with number responding, the mean and the standard deviation across individual ratings. For the first three variables, the highest rating was a '5' and the lowest '1'. The range on the last variable was '3' to '1' representing the values of positive, neutral or negative. The reason for the three point scale was that it was based on similar items constructed by one of the authors in an unrelated investigation where results yielded acceptable reliability.

The first item for Organizational Actions, 'obtain knowledge of external resources', was the only item to receive a rating above the midpoint. The subscale for Exercise Outcomes yielded results associated with interaction and planning with other agencies both within the respondent's business sector and external to that sector. For Exercise Design Features, it can be noted (although not displayed here) that many of the items received fairly high ratings of 4.0 and above indicating satisfaction with many of the features. Respondents considered the 'organization of the exercise' and 'group decision-making' to be particularly effective. The fourth subscale of the survey, Organizational / Personal Characteristics, was designed to primarily capture the 'organizational culture' of the respondent. The items included both personal and organizational elements that have been shown to influence the application of new learning. The top-rated item was 'having the correct background for the exercise' which may be translated as making the sure the right people are attending the training session. 'Opportunities to share or apply what was learned' were the next two highest ratings. The last two items were associated with 'barriers to apply new learning' at the individual and organizational level. These items achieved a rating of essentially 'neutral' possibly meaning that such barriers can exist in an organization, although it not considered as a strong negative influence.

Subscales

Much of the remainder of the analysis will examine the use of subscales to represent the four primary variables of the survey. The descriptive data for these subscales is shown in Table 2. A subscale was constructed by the simple addition of the individual item ratings in its group. For example, there are 11 items in the Organizational Actions variable rated on a scale from 1 to 5. Adding across the items, the maximum "subscale" score would be 55 and the lowest 11. In Table 2,

Table 1
Top-Ranked Items for Four Subscales (by Mean Rating)

Subscale / Item	N	Mean	Std. Deviation
Organizational Actions (11 items; 5-point scale)			
Obtained first-hand knowledge of external resources available outside my organization/agency	44	3.29	1.19
Established regular meeting schedule with other organizations	44	2.38	1.41
Created/maintained mutual aid agreements with other organizations or agencies	44	2.18	1.16
Exercise Outcomes (15 items; 5-point scale)			
Interaction with other organizations w/in same industry or government sectors	47	4.12	0.76
Collaborative planning with wide range of agencies	46	3.97	0.93
Interaction with other organizations in different industry or government sectors	47	3.93	0.96
Exercise Design Features (10 items; 5-point scale)			
Organization of exercise activities	40	4.20	0.75
Group decision-making process	41	4.14	0.88
Organizational / Personal Characteristics (11 items; 3-point scale)			
Having the correct background (skills) for role in exercise	41	2.92	0.26
Opportunity to share what was learned with colleagues	42	2.71	0.55
Opportunity to use what was learned in exercise	40	2.70	0.46
At the organizational level, there exist barriers to applying new learning	36	2.00	0.67
At the employee level, there exist barriers to applying new learning	36	1.94	0.62

the minimum score obtained was 11 (meaning at least one respondent gave '1's to each item) and maximum was 46, considerably below the maximum possible. The standard deviations represent the amount of variability, and subscales with more items should show more variability. Since the last subscale on Personal / Organizational Characteristics has 11 items and only a 3-point scale, its standard deviation is the lowest as might be expected. Following the label in each row is the Coefficient alpha reliability measure for each scale. With the lowest of these at 0.84 and the highest at 0.93, all values indicated considerable subscale internal consistency. The subscales were created via content and face validity from existing literature and the considerable experience of one of the authors; there was not sufficient sample size to group items from a factor analysis and no items were dropped from the subscales. Such modifications are likely in future use, however.

Table 2
Subscale Descriptive Statistics

Subscale / (Reliability Coefficient)	No. of Items	Sample Size	Minimum Obtained	Maximum Obtained	Mean	Standard Deviation
Organizational Actions / (0.88)	11	44	11.0	46.0	29.73	9.89
Exercise Outcomes / (0.92)	15	47	31.0	74.0	55.38	11.07
Exercise Design Features / (0.93)	10	41	24.0	50.0	39.27	7.50
Personal / Organizational Characteristics / (0.84)	11	42	13.0	33.0	25.83	4.22

An intercorrelation matrix was developed to learn of the inter-relationships among the subscales. The results were as follows:

- Using Organizational Actions as the proxy for Lessons Learned, the other three scales were significantly related to the criterion (0.39, $p=.008$; 0.41, $p=.006$; 0.42, $p=.008$) respectively for Outcomes, Design Features and Characteristics. This finding offers important support for the logic model.
- The highest relationship was between Exercise Outcomes and Exercise Design Features ($r= 0.70$, $p<.001$), thus they overlap in measurement with shared variance around 50%.
- Personal / Organizational Characteristics appeared to be related only to Organizational Actions ($r = 0.42$, $p=.006$) but not the other two (Exercise Outcomes, $r = -0.09$, $p= .594$; Exercise Design, $r = 0.19$, $p= .252$).

Organizational Actions Analysis. The proposed logic model uses Organizational Actions as a surrogate measure for actual response or preparation effectiveness since the latter can be measured by only observation or evaluation of field performance. Thus, the remainder of this analysis uses that subscale as the criterion to examine other findings.

The first analysis examines whether or not any of the demographic variables have an influence on Organizational Actions. Since the demographic variables are nominal, t-tests were performed on each of the variables. All results were not significant with the exception that those respondents with limited experience implemented fewer actions ($t = 3.2$, $p = 0.043$). In a practical way, this finding indicates that virtually all types of respondents can demonstrate impact of lessons learned from an RBX™ exercise, as measured by the Organizational Actions scale, with the possible exception for those with limited experience.

The next set of analyses examined what elements in the other three subscales had the most relative influence on Organizational Actions. In other words, which items in the various scales seemed most likely to lead to organizational action? To perform this analysis, the Organizational Actions subscale was correlated with each item from the other three subscales. All results with significant statistical relationships ($p < 0.05$) are listed in Table 3.

Table 3
Significant Organizational Actions Subscale Correlations with Other Items

Subscale / Item	r value	p value
Exercise Outcomes		
• Need for understanding of own organizational responsibilities	0.44	.003
• Collaborative planning with wide range of agencies	0.42	.005
• Understanding of the role of media	0.34	.022
• Testing of disaster / emergency plans	0.34	.025
• Increased relationship with local first responder organizations	0.32	.032
Exercise Design Features		
• Relevance of scenario to personal job responsibilities	0.51	.001
• Accomplishment of exercise objectives	0.38	.014
• Quality of interaction with other participants	0.38	.017
• Evaluation of actions/decisions across groups	0.38	.015
• Group decision-making process	0.35	.026
Organizational Culture & Personal Characteristics		
• Opportunity to share what was learned with colleagues	0.63	<.001
• Match of exercise objectives to current job requirements	0.58	<.001
• Opportunity to use what was learned in exercise	0.56	<.001
• Immediate manager supports application of new learning	0.38	.014
• Organization supports application of new learning to job duties	0.33	.036

A final analysis examined the extent to which the Organizational Actions subscale could be predicted from all other variables by employing multiple regression analysis. That analysis showed that the other three subscales (Exercise Outcomes, Exercise Design Features and Organizational/Personal Characteristics) were significant predictors as well as one of the demographic variables, Type of Organization. The total percent of variance accounted for in the criterion was approximately 59%. This value was considered as an exceptionally high result in comparison to other results in the social science domain, especially for an evaluation tool that has yet to be refined.

Results of Qualitative Analysis from the Online Survey. There were four open-ended questions on the online survey. These questions were generated from the first round of interviews. Each question was subject to a constant comparison method based on question, participant, and region of exercise. The final constant comparison produced no new themes and reached a point of redundancy. The response numbers vary for each question as some participants did not answer all of the open-ended questions.

How did this exercise differ from other exercises in which you have participated? While answers to this question varied, a couple of common themes emerged. Eleven respondents indicated that this exercise contained a more multi-organizational focus. Such responses included "[This exercise] was designed specifically to highlight the intersection of government and industry roles," "We have not participated in an exercise with different government agencies," and, "… inclusion of government and industry was very helpful." Similar to the notion of multi-organizational participation, seven respondents commented on the interactive design of the exercise. The interactive design element was portrayed in responses such as, "Better interaction," "More interactive," and "[This exercise had a] greater level of interactivity and opportunity to engage a wide range of disciplines and backgrounds." The final common theme, of which six respondents agreed, was the "real-ness" of the design. The real-life theme was described by respondents as "It was like a real life scenario that could take place," "The exercises were very realistic and you had to remember that it was an 'exercise'," and "[The exercise] mirrored real events in many ways."

How were lessons learned from the exercise communicated with your organization? This question produced two common themes. A quarter of respondents, eight, portrayed that lessons learned from the exercise were communicated verbally such as in meetings or discussion. Responses that depicted verbal communication included, "Verbally," "Staff meeting discussions," and "Verbal reports to all involved." Conversely to verbal representations of communicating lessons learned, six of the 32 respondents indicated that they distributed the After Action Report.

Are there things that your organization does to facilitate or inhibit the lessons learned from such exercises? The analysis of this question did not provide a clear common theme. Answers included a wide range of topics such as bureaucratic composition of the organization, to lack of organizational motivation, to a lack of organizational communication. The only topic repeated was an organizations' ability to inhibit the lessons learned process due to budgetary constraints. Three of the 28 respondents commented on organizational budget or money confines.

What can your organization do to employ the lessons learned from the exercise? There was only one common theme to report albeit this question produced more significant responses with respect to the theme. Of the 29 responses, 8 of the respondents indicated that their organization can use training, exercise or seminars to employ the lessons learned from the exercise. Common responses to this questions included, "Increased training for staff," "Annual exercises," "Need internal exercises that involve more participants," and "Conduct additional discussion based exercises on this subject with different goals and objectives, moving into a tabletop exercise or functional exercise."

SUMMARY

The general conclusion from this initial validation study of the tool used to measure the long-term effectiveness of the discussion-based exercises of the type investigated was that it offered very promising results. Revisiting the first research question, the logic model does appear to suggest the ability to predict organizational action. The initial confirmation of most of the logic model, the high power for predicting organizational actions, the high reliability of the four subscales, and the practical significance associated with identification of key item relationships with the organizational actions criterion all offer strong evidence for the future investigations using this tool.

Five sets of variables were proposed within the logic model to depict relationships of the lessons learned process. One of those variables, Organizational Actions, was used as a surrogate for field-based measure of coordinated response / preparation effectiveness. Thus, the major hypothesis of the study was that Organizational Actions could be related (that is, predicted) from the other four variable sets. This hypothesis was confirmed on a preliminary basis from the findings. Specifically, the variable sets (subscales) of Exercise Outcomes, Exercise Design Features, Organizational/Personal Characteristics, and one demographic variable (industry sector) were shown to share 59% common variance with the Organizational Actions subscale. This strong statistical relationship offers sound evidence to refine the use of this evaluation tool.

In considering the third research question, there is some practical significance worthy of exploration associated with identifying key items related with the Organizational Action criterion. Organizational Actions was meant to represent the deployment of "lessons learned", that is, initiatives, strategies, and improvements that organizations would undertake to enhance coordination for emergency and disaster response effectiveness. Since they represent actual changes implemented, it is not surprising that we found relatively low mean ratings (all except one were below the midpoint of the 5-point scale). One could also interpret that the top two of the 11 items were associated with organizing internal plans (assets available from other agencies and defining internal responsibilities). Conversely the lowest two ranked items were associated with involvement of external agencies (establishing a regular meeting schedule and crafting mutual aid agreements), which would likely be more difficult to sustain.

The 15 item means representing Exercise Outcomes supported the general intent of the RBX™. In other words, the top three rated items were all concerned with cross-agency discussions and planning, key elements needed to support coordinated disaster response (see Drabek, 2005). The subscale created for Exercise Outcomes was significantly related to the 10-item subscale on Exercise Design Features. This is an expected finding in that if the exercise was designed to promote interactions that would lead to coordination, then discussion activities should be part of the exercise design. Thus, appropriate design of the learning activities should promote the immediate outcomes of the exercise, and the two factors should have a strong relationship. Other exercises where objectives and learning activities do not match as well as compared to our findings would most likely not show this strength of relationship.

Another expected finding was the very low relationship between the subscale representing organizational culture, Organizational/Personal Characteristics, and Exercise Outcomes and Exercise Design Features, and yet a statistically significant relationship with Organizational Actions. The organizational culture variable represented factors outside the exercise event and thus, not part of what could be built into a given exercise. The only possible exception to this was the highest rated item on having the right background for exercise participation. This implies that the 'right people' need to be the ones selected for exercise participation if subsequent action in the organization is to occur. The general low relationship of organizational culture to other variables except subsequent action demonstrates its unique value to understanding how lessons can actually get learned or transferred. More specifically, organizations and their leaders behave in ways that can support or thwart the development of change that relate to improvements in organizational response capability.

When the Organizational Actions subscale was used as a criterion measure to examine key items within each of the other subscales, some patterns in the analyses suggest the following conclusions. Subsequent actions or changes to improve response effectiveness are more likely if the exercise is customized to the audience. This requires tailoring the exercise to job responsibilities and also requires participants to be familiar with their organization's role in emergency / disaster response. In addition to those subsequent changes to improve response effectiveness, actions that lead to coordination are related to how the exercise is designed and conducted. An emphasis on meaningful interaction across groups, group decision-making, and concentration on accomplishing those objectives in an organized fashion appear warranted. Finally, an understanding of how lessons can be translated into action requires that participants acquire an understanding of their own organizational culture. Lessons can get learned if people take the time to discuss implications of their exercise participation and are supported in testing out this learning to address internal response capabilities.

CONCLUSIONS

In one of the first of its kind, this study examined longitudinally key variables hypothesized to influence the lessons learned process. These variables then formed the basis for the logic model and the assessment tool in order to measure long term effectiveness as it pertains to discussion-based exercises.

For future research considerations, there are a number of limitations that should be addressed to further increase its reliability and validity. First, from a methodological perspective, this research was not a true longitudinal study, but rather a longitudinal perspective on effectiveness. That is, a true longitudinal study would require tracking information from the exercise conduct to subsequent action by matching respondents from beginning to end so that changes over time could be documented. Our study collected data at a single point in time and therefore, trusted the accurate memories of the respondents. Further, the time lapse of more than one year is too long for participants to recall information associated with one particular exercise. This is especially true for individuals who yearly participate in many exercises. Second, the study needed a larger sample size to refine the tool and perform item analysis. An item analysis will allow us to better understand the observed relationships and make more targeted conclusions concerning its utility and limitations. A larger sample would have permitted inter-item correlations, a factor analysis, and more detailed reliability analysis for the subscales. The relatively high subscale reliability values, while a very positive outcome, may also mean that there is considerable item redundancy. Such redundancy could be reduced by selecting items that provide the highest discrimination between predictor variables and the criterion. Finally, this analysis should shorten the length of the tool so that we might be able to counteract respondent fatigue (complete responses to items tended to trail off near to the end of the survey) and provide greater ease of administration.

This tool provides a means to a longitudinal perspective on the potential effectiveness of emergency and disaster preparedness exercises where a major emphasis is on cross-agency coordination. While the subject of this investigation was a special discussion-based exercise for food safety, it is expected that the tool can be applied to other discussion-based formats. Specifically, the use of this tool could be expanded to other emergency management areas outside food safety with minor adjustments to some of the items. These adjustments should not materially impact the underlying relationships found in the current investigation.

REFERENCES

Auf der Heide, E. (1989). *Disaster response: Principles and Preparation and Coordination.* St. Louis, MO: The C.V. Mosby Company.

Carley, K. M., & Harrald, J. R. (1997). Organizational learning under fire. *American Behavioral Scientist, 40,* 310-332.

Drabek, T. E. (2005). Predicting disaster response effectiveness. *International Journal of Mass Emergencies and Disasters, 23* (1), 49 - 72.

Drabek, T. E. & McEntire, D. A. (2002). Emergent phenomena and multi-organizational coordination in disasters: Lessons from the research literature. *International Journal of Mass Emergencies and Disasters, 20* (2), 197 – 224.

Drabek, T. E. (2001). Coordinating disaster responses: A strategic perspective. *The American Society of Professional Emergency Planners, 8*, 29-40.

Drabek, T. E. (1994). *Disaster evacuation and the tourist industry.* Institute of Behavioral Sciences, University of Colorado, Boulder, CO.

Drabek, T. E. (1985). Managing the emergency response. *Public Administration Review, 45*, 85-92.

Gillespie, D. F., & Streeter, C. L. (1987). Conceptualizing and measuring disaster preparedness. *International Journal of Mass Emergencies and Disasters, 5*(2), 155-176.

Gillespie, D. F., & Banerjee, M. M. (1993). Prevention planning and disaster preparedness. *The Journal of Applied Social Sciences, 17*(2), 237-253.

Holton, E. F. III (2005). Holton's evaluation model: New evidence and construct elaborations, *Advances in Developing Human Resources*, *7* (1), 37 – 54.

Kinsey, J. (2007). Best practices for defense against terrorism: What consumers expect food companies to do. Downloaded from the National Center for Food Protection and Defense website (http://www.ncfpd.umn.edu/), November 30, 2007).

Mileti, D. S., Darlington, J. D., Fitzpatrick, C., & O'Brien, P. W. (1993). *Communicating earthquake risk: Societal response to revised probabilities in the Bay Area.* Hazards Assessments Lab and Department of Sociology, Colorado State University, Fort Collins, CO.

Paton, D., Johnston, D., & Houghton, B. F. (1998). Organizational response to volcanic eruption. *Disaster Prevention and Management, 7*, 5-13.

Petak, W. J. (1985). Emergency management: A challenge for public administration. *Public Administration Review, 45*, 3-7.

Potter, M.A., Ley, C. E., Fertman, C. I., Eggleston, M. M., & Duman, S. (2003). Evaluating workforce development: Perspectives, processes, and lessons learned. *Journal of Public Health Management Practice, 9*(6), 489-495.

Quarantelli, E. L. (1997). Ten criteria for evaluating the management of community disasters. *Disasters, 21*, 39-56.

Tierney, K. J. (1994). Proceedings from NEHRP Conference and Workshop on Northridge, CA Earthquake of January 14, 1994. *Research overview: Emergency response*.

Toulmin, L. M., Bivans, C. J., & Steel, D. L. (1989). The impact of intergovernmental distance on disaster communications. *International Journal of Mass Emergencies and Disasters, 7,* 116-132.

US Department of Agriculture (2005), Food Safety and Inspection Service. Industry self-assessment checklist for food security. Downloaded from www.fsis.usda.gov, April 15, 2007.

Waugh, W. L. (1993). Coordination or control: Organizational design and the emergency management function. *Disaster Prevention and Management, 2*, 17-31.

Webb, G.R., Tierney, K.J., & Dahlhamer, J.M. (2000). Businesses and disasters: Empirical patterns and unanswered questions. *Natural Hazards Review, 1* (2), 83-90.

Wenger, D. E. (1989). The study of emergent volunteer and emergent organization response in search and rescue: Approaches and issues for future research. *HRRC Publication 5P.* College Station, TX: Hazard Reduction and Recovery Center, Texas A&M University.