DOI: 10.1111/j.1539-6924.2012.01837.x

Context-Specific, Scenario-Based Risk Scales

Michael Yu,[1,*] Tomás Lejarraga,[2] and Cleotilde Gonzalez[1]

Reacting to an emergency requires quick decisions under stressful and dynamic conditions. To react effectively, responders need to know the right actions to take given the risks posed by the emergency. While existing research on risk scales focuses primarily on decision making in static environments with known risks, these scales may be inappropriate for conditions where the decision maker's time and mental resources are limited and may be infeasible if the actual risk probabilities are unknown. In this article, we propose a method to develop context-specific, scenario-based risk scales designed for emergency response training. Emergency *scenarios* are used as scale points, reducing our dependence on known probabilities; these are drawn from the targeted emergency *context*, reducing the mental resources required to interpret the scale. The scale is developed by asking trainers/trainees to rank order a range of risk scenarios and then aggregating these orderings using a Kemeny ranking. We propose measures to assess this aggregated scale's internal consistency, reliability, and validity, and we discuss how to use the scale effectively. We demonstrate our process by developing a risk scale for subsurface coal mine emergencies and test the reliability of the scale by repeating the process, with some methodological variations, several months later.

KEY WORDS: Emergency response; mining; risk communication; scale development; scale evaluation

1. INTRODUCTION

Six months after the Sago Mine disaster, the Mine Improvement and New Emergency Response Act of 2006 was signed into law to improve mine emergency response. One consequence of the act was the rapid development of refuge alternatives—temporary shelters located inside the mine for when escape is infeasible. As successfully escaping the mine is safer than remaining in these shelters, refuge alternatives are described to miners as being a "last resort." Anecdotal evidence from mine safety researchers suggests that when asked about refuge alternatives, many miners do not initially view them as alternatives at all, claiming that they would never use them under any circumstance. Only after being pressed further do some of the miners change their minds, acknowledging that refuge alternatives may be appropriate under certain conditions.

Ambivalence regarding the use of refuge alternatives could be explained by the imprecise guidance provided on when they should be used. Training materials are specific on when not to use refuge alternatives, (e.g., "Smoke should never prevent escape"), but vague on when to use them (e.g., "Only as a last resort").[1,2] Clearer communications might promote more stable and consistent views on refuge alternative use, better coordination among mining crews, clearer expectations for rescue operators, and more appropriate use of refuge alternatives. A mine emergency risk scale, if developed, could be used as part of the communication of when refuge alternatives should be considered. Given the stressful conditions

[1]Dynamic Decision Making Laboratory, Carnegie Mellon University, Pittsburgh, PA, USA.
[2]Departament d'Economia de l'Empresa, Universitat de les Illes Balears, Palma de Mallorca, Spain.
*Address correspondence to Michael Yu, Dynamic Decision Making Laboratory, Carnegie Mellon University, 4609 Winthrop Street, Pittsburgh, PA 15213, USA; msyu@andrew.cmu.edu.

0272-4332/12/0100-0001$22.00/1

under which it would be needed, the scale should be easy to interpret, adaptable to dynamic and novel risks, and consistently interpreted by miners.

The focus of this article is to present a methodology to create context-specific, scenario-based risk scales for emergency response training through the rankings of emergency scenarios aggregated from a sample of the target training population. We describe how to develop the scale, propose measures to evaluate it, and consider how to use it effectively. As an example of this process, we describe two surveys conducted with employees from the National Institute for Occupational Safety and Health (NIOSH) and subsurface coal mine inspectors, used to develop the methodology and identify potential improvements.

2. EMERGENCY RESPONSE AND RISK SCALES

Effective emergency decision-making demands that the decision maker identify the type of environment he or she is in and then take the appropriate response. However, the nature of emergencies makes these decisions difficult, with a constantly changing environment and limited time to make decisions. Errors in behavior could arise from inaccurate assessments of the environment or inappropriately trained responses. Under a signal detection theory framework, these can be attributed to low sensitivity and an inappropriate response bias, respectively.

Most current research on risk scales focuses on how to communicate accurate probability estimates of static risks for deliberative decision making, e.g., in medical informed consent,[3,4] nuclear waste management,[5] or electromagnetic radiation.[6] Such scales do not meet the needs of emergency response in at least two ways. First, accurate probability estimates posed by an emergency are often unavailable, making the development of such scales infeasible. Second, these scales assume that decision makers have the resources to translate the risk they are facing onto some other measure of risk, such as a numeric probability scale or common everyday risks. This translation may introduce noise in the assessment, which is exacerbated by limited time and mental resources. In turn, this may lead to greater errors in risk assessments.

Of the existing risk scales, probability-based risk scales are often used as they are easy to analyze quantitatively. However, people have been shown to interpret and produce numeric probabilities inconsistently; often due to heuristics, such as anchoring[7] or emotions,[8] and due to individual differences.[9] Frequency estimates have been proposed as more intuitive,[10–12] but further research suggests this is not always the case.[13,14] To reduce issues in interpretation, some researchers advocate providing familiar risks along with numeric risk scales; for example, calibrating household radon risks to smoking a certain number of cigarettes each day.[15] Although familiar risks can reduce interpretation concerns, this presentation has been criticized based on the differing natures of the various risks involved.[16] For example, smoking is voluntary, costly to consume, but provides pleasure to the smoker; whereas radon is involuntary, costly to remediate, and provides no benefit to the homeowner.

2.1. Context-Specific, Scenario-Based Risk Scale

For emergency response training, developing a risk scale using scenarios drawn from the emergency context may be more effective than using unrelated but familiar risks. First, this allows the scale to be developed without reference to known probabilities. Second, this may reduce the mental resources required to locate an actual risk on the scale by reducing contextual differences. To this end, we propose a context-specific, scenario-based risk scale developed by aggregating the same risks rank ordered independently by the people who will need to use the scale. A summary of the scale development process is provided in Fig. 1 and described in further detail below.

2.1.1. Developing the Scale

The first step to developing the scale is to define its purpose. This provides guidance on the types of scenarios to include. For example, as our task was focused on developing a scale to train emergency response for underground coal miners, we designed our scenarios to address immediate risks (structural collapse, explosions) rather than progressive risks (exposure to coal dust).

The next step is to define the endpoints and precision of the scale. Scale endpoints should cover the range of risks that the target population would face. In our case, we used scenarios that we expected to be clearly benign (stubbed toe) to clearly dangerous (explosion toward the mine exit). For scale precision, we included more scenarios where the right emergency response was uncertain and fewer where the response seemed evident.

Development	Validation Tests
Define purpose	Internal consistency (Kendall's W, Pairwise Agreement)
Determine scale end-points and scale precision	Scale Reliability (Kendall's tau)
Develop emergency risk scenarios for scale	Scale Validity (Kendall's tau)
Survey target population for scale rank-ordering	
Aggregate the scale	

Fig. 1. Scale development process.

Finally, a sample from the target population is asked to rank order the scenarios by level of risk. Then, the rank orderings are aggregated into a single risk scale. We propose using the Kemeny–Young method, described later, to aggregate the scales for two reasons. First, the Kemeny–Young method can be interpreted as estimating a "true" rank ordering based on individual samples with noise.[17] This interpretation is consistent if we believe that our training population is able to judge the relative risk of our scenarios, even when these judgments are potentially imperfect. Second, the Kemeny ranking is easy to explain as maximizing the number of people who would agree with the rank ordering of any two risks selected at random from the aggregate scale. A simple interpretation may be important in building credibility if the process needs to be communicated to or agreed upon by scale users.

Deriving the Kemeny ranking is simple though computationally intensive. First, determine all possible rank orderings. For each rank ordering, identify all possible pairs of scenarios and calculate the percentage of respondents who agree with the rank ordering of each pair. Average the pairwise agreement for each rank ordering. Finally, select the rank ordering with the highest average pairwise agreement. A pairwise agreement matrix (see Fig. 3 for example), which is a table of the risk scenarios included as both column and row headings and the sample agreement to their rank orderings provided as cell values, can be used to manually identify the Kemeny ranking without processing through all possible rank orderings. The researcher would simply rearrange the rows and columns, making sure that the row and column labels are symmetric, to maximize the percentages in the upper right diagonal of the matrix. For more details on the methodology, we recommend Young's work[17] and for a comparison with alternate models, we recommend the summary of vote-counting processes provided by Levin and Nalebuff.[18]

2.1.2. *Validating the Scale*

After developing the aggregated risk scale, we validate it in several ways. First, we evaluate internal consistency, which we separate into (1) how well respondents agree with each other and (2) how well respondents agree with the aggregated scale.

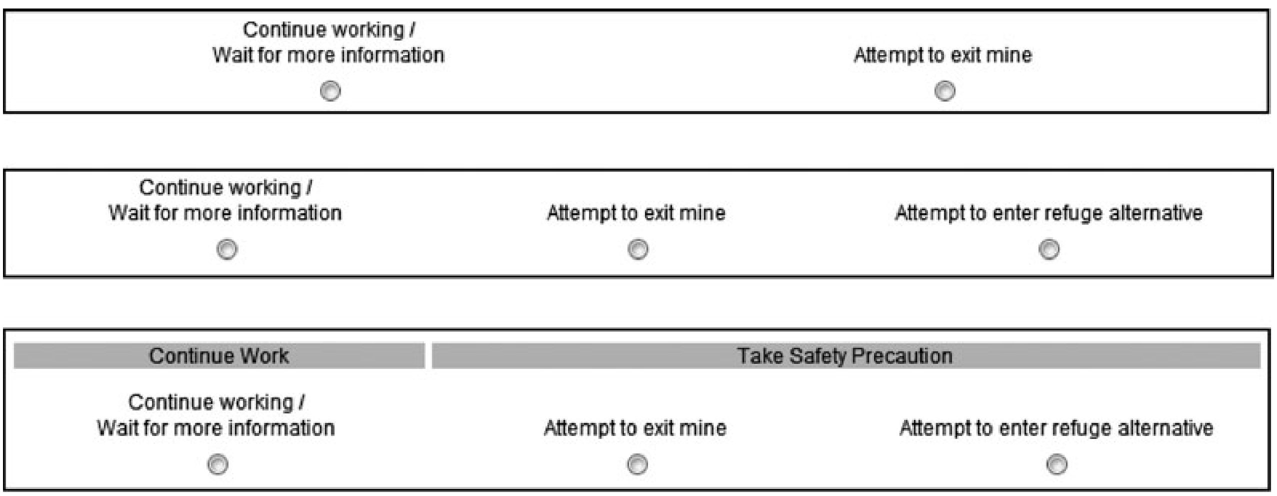

Fig. 2. Response alternative formats.

	Hurt foot	Temporary haze	Break in ventilation	Burnt-rubber smell	Roof fall along the face	Roof fall outby	Reverse in ventilation	Stop in ventilation	Elevated level of methane	Smoke outby	Explosion and hurt shin	Explosion outby
1. Hurt foot		72%	83%	78%	94%	100%	100%	100%	100%	100%	100%	100%
2. Temporary haze	28%		56%	61%	83%	78%	78%	100%	94%	100%	100%	100%
3. Break in ventilation	17%	44%		56%	72%	78%	94%	100%	94%	89%	83%	94%
4. Burnt-rubber smell	22%	39%	44%		78%	78%	78%	100%	94%	100%	100%	100%
5. Roof fall along the face	6%	17%	28%	22%		56%	56%	50%	83%	89%	83%	100%
6-7. Roof fall outby	0%	22%	22%	22%	44%		50%	56%	83%	89%	89%	100%
6-7. Reverse in ventilation	0%	22%	6%	22%	44%	50%		67%	78%	78%	78%	89%
8. Stop in ventilation	0%	0%	0%	0%	50%	44%	33%		61%	72%	72%	78%
9. Elevated level of methane	0%	6%	6%	6%	17%	17%	22%	39%		78%	78%	78%
10. Smoke outby	0%	0%	11%	0%	11%	11%	22%	28%	22%		78%	83%
11. Explosion and hurt shin	0%	0%	17%	0%	17%	11%	22%	28%	22%	22%		56%
12. Explosion outby	0%	0%	6%	0%	0%	0%	11%	22%	22%	17%	44%	

	Hurt foot	Break in ventilation	Burnt-rubber smell	Roof fall along the face	Pillar squeezing	Roof fall outby	Reverse in ventilation	Mantrip derailing and fire	Stop in ventilation	Elevated level of methane	Smoke outby	Explosion and hurt shin	Explosion outby	Explosion and broken leg	Explosion and methane
1. Hurt foot		61%	71%	79%	84%	79%	77%	82%	82%	82%	93%	89%	91%	89%	89%
2. Break in ventilation	39%		66%	70%	63%	68%	73%	79%	86%	80%	88%	91%	88%	89%	84%
3. Burnt-rubber smell	29%	34%		59%	54%	59%	71%	68%	77%	70%	86%	86%	89%	91%	91%
4. Roof fall along the face	21%	30%	41%		52%	52%	57%	64%	61%	59%	77%	82%	86%	86%	88%
5. Roof fall outby	16%	38%	46%	48%		52%	59%	63%	75%	73%	86%	86%	91%	91%	91%
6. Pillar squeezing	21%	32%	41%	48%	48%		55%	61%	64%	68%	84%	80%	84%	91%	89%
7. Reverse in ventilation	23%	27%	29%	43%	41%	45%		61%	55%	64%	80%	84%	89%	86%	89%
8. Mantrip derailing and fire	18%	21%	32%	36%	38%	39%	39%		61%	61%	77%	80%	86%	88%	88%
9. Stop in ventilation	18%	14%	23%	39%	25%	36%	45%	39%		63%	77%	82%	82%	88%	84%
10. Elevated level of methane	18%	20%	30%	41%	27%	32%	36%	39%	38%		79%	82%	88%	86%	88%
11. Smoke outby	7%	13%	14%	23%	14%	16%	20%	23%	23%	21%		59%	70%	68%	79%
12. Explosion and hurt shin	11%	9%	14%	18%	14%	20%	16%	20%	18%	18%	41%		63%	79%	68%
13. Explosion outby	9%	13%	11%	14%	9%	16%	11%	14%	18%	13%	30%	38%		54%	73%
14. Explosion and broken leg	11%	11%	9%	14%	9%	9%	14%	13%	13%	14%	32%	21%	46%		63%
15. Explosion and methane	11%	16%	9%	13%	9%	11%	11%	13%	16%	13%	21%	32%	27%	38%	

Fig. 3. Pairwise agreement matrices. Top: NIOSH employees; Bottom: Mine inspectors.

Respondent agreement with each other can be measured using Kendall's W, a nonparametric measure of agreement in rank orderings that ranges from 0 (no pattern) to 1 (perfect agreement) and is interpreted similarly to an average correlation coefficient. To our knowledge, there is no standardized measure of individual agreement to an aggregate scale, so we propose the average pairwise agreement. For a Kemeny ranking, this ranges from 50% (no pattern) to 100% (perfect agreement). There are two cautions in interpreting these measures. First, both are sensitive to the scenarios chosen, such that the more similar the risk of the scenarios included in the scale, the more likely a disagreement in their rankings would occur. Including scenarios of similar risk thus lowers our measures of internal consistency, yet offers greater precision and understanding of the population's perception of the risks. As such, lower values do not necessarily represent a low-quality scale. Second, scale-level measures may obscure important disagreements or uncertainty about the risk rankings of specific scenario pairs. Thus, the full pairwise agreement matrix is still necessary to comprehensively evaluate internal consistency.

Scale reliability provides us with a measure of how consistently the scale is interpreted across groups, such as different subsets of the population or the same population over time. We propose Kendall's tau, a nonparametric measure used to assess agreement in two rank orderings, and Kendall's tau-b, which provides a similar measure accounting for ties. Kendall's tau and tau-b range from –1 (perfect disagreement) to 1 (perfect agreement).

Scale validity provides us a measure of whether our scale is actually measuring what we believe it is measuring, and can be evaluated by comparing the developed scale with another rank ordering of the same scenarios by risk. Ideally, if quantitative risk estimates are available, we can compare the quantitative rank-ordering of the scenarios with our aggregated rank orderings using Kendall's tau/tau-b. Where the quantitative estimates do not exist, we look to compare how our aggregated rank orderings align with an alternate method to order the scenarios. In our case, we use reported emergency response behavior as a second proxy for risk.

3. MINE EMERGENCY RISK SCALE, SURVEY 1

To evaluate the proposed methodology, we worked with NIOSH to develop a scenario-based risk scale for use in training emergency mine response. In consultation with NIOSH, 12 scenarios were originally developed and piloted with NIOSH employees. After the pilot study, one scenario was removed and four scenarios were added to improve precision around risks and risk levels that were of greater interest in mine safety communication. In addition, as part of our collaboration with NIOSH, a secondary study was included in the scale-development survey, which asked how the introduction and presentation of refuge alternatives might affect emergency response decisions. We discuss the secondary study later, as data from the study are used as part of our validation of the scenario-based risk scale and as the study provides an example of how the scale may be used.

3.1. Methodology

3.1.1. Participants

Eighteen NIOSH employees volunteered to pilot the initial survey and provided feedback. Employee backgrounds included mine safety researchers, mining engineers, computer engineers, and behavioral scientists. Respondents had worked an average of 7.1 years ($SD = 10.4$) in the mining industry, of which four respondents reported years worked inside a mine ($M = 5.3$, $SD = 3.3$). Average age of respondents was 38.7 years ($SD = 11.3$), and 38.8% of respondents were female.

Sixty-two mine inspectors were asked to complete the revised survey. Based on demographic responses, one participant who reported being born in 1900 and reported no experience working in the mining industry was dropped from analysis. Of the remaining 61 mine inspectors, respondents worked an average of 29.6 years ($SD = 9.7$) in the mining industry and all reported some years of working inside a mine ($M = 25.5$, $SD = 10.5$). Average age of respondents was 50.9 years ($SD = 8.67$), and none of the respondents were female.

3.1.2. Materials

Respondents completed a survey that was administered online. The survey asked respondents how they would respond to a given mine emergency scenario, to rank order the scenarios, and then to provide demographic information about themselves.

The survey was introduced with the following text:

> On the following page, you will be presented with several scenarios that involve some level of risk. We ask that you consider these scenarios as if you were an underground coal miner.
>
> Imagine that you just started a shift and are currently at the mine face, approximately 40 crosscuts inby the entrance. A refuge alternative is located 10 crosscuts outby your position. Please take a moment to picture yourself in this situation.
>
> Now, read the following scenarios and tell us which action you feel would be most appropriate. While real-world scenarios may be more complex, we ask you to provide the best answer given the question.

Respondents were then presented with different mine risk scenarios in randomized order and were asked to indicate how they would respond given a set of two to three response alternatives including: "Continue working/Wait for more information," "Attempt to exit mine," and "Attempt to enter refuge alternative." The response alternatives were presented in three different formats, with each participant viewing only one format. In the 2-Flat (2F) format, only the options to continue work or exit the mine were presented. In the 3-Flat (3F) format, all three response alternatives were presented in the same manner. In the 3-Collapsed (3C) format, all options were presented, however the options to exit the mine or enter the refuge alternative appeared under the "Take Safety Precaution" subheading and the continue work option appeared under the "Continue Work" subheading. The three presentation formats are depicted in Fig. 2.

The mine risk scenarios presented to respondents are included in Table I. After indicating how they would respond to each of the scenarios, respondents were presented with the full list of scenarios in a randomized order and asked to rank order the scenarios from "most risky" to "least risky" using a drag-and-drop interface.

3.1.3. Design

Given slightly different scenarios, a separate risk scale was developed for NIOSH and the mine inspectors. Internal consistency and validity measures were calculated for each scale, whereas reliability

Table I. Mine Risk Scenarios

Risk Scenarios	NIOSH	Mine Inspectors
Hurt foot. As you were working, you took an incorrect step and hurt your foot. There is a sharp, minor pain that quickly subsides. You are able to walk and move without issue.	X	X
Temporary haze. As you are working you notice that it seems a bit harder to see, and you notice large dark particles drifting in the air. As you continue working, these particles seem to dissipate.	X	
Break in ventilation. While working you notice the air quantity in your entry has dropped significantly. After a minute, you notice that ventilation is restored.	X	X
Burnt-rubber smell. As you work, you smell something that reminds you of burnt rubber. As you continue working, the smell seems to go away.	X	X
Roof fall along the face. You hear a loud disturbance coming from a location further down the mine face. A call from miners mine in that location indicates that an area of roof near them has collapsed and weakened the adjoining roof.	X	X
Pillar squeezing. You notice three pillars in the area adjacent to you showing signs that they are beginning to crush from the roof load. You know from the mine maps that this area trends toward deeper cover.		X
Roof fall outby. You hear a loud disturbance coming from outby. A call from miners who were just entering the mine indicates that they believe an area of roof has collapsed and weakened the adjoining roof.	X	X
Reverse in ventilation. While working you notice that the air in your entry has changed direction.	X	X
Mantrip derailing and fire. Soon after you arrive, you hear a loud disturbance outby. A call from miners in the area indicates that the mantrip derailed on its way back up, damaging the track and causing an electrical fire. The miners in the area believe that they have the fire under control.		X
Stop in ventilation. While working you notice the air quantity in your entry has dropped significantly. After 10 minutes, you find that ventilation has still not been restored.	X	X
Elevated level of methane. When methane levels in your area reached 1% you made changes to the local ventilation to try and correct the problem. Methane levels have reached 1.3% and are still climbing.	X	X
Smoke outby. You notice heavy dark smoke coming up the entry you are in from the outby direction. You are not aware of any activities going on in the mine that could cause the smoke, and, as time passes, the smoke appears to get thicker.	X	X
Explosion and hurt shin. There was an explosion in the mine coming several crosscuts away. The explosion dislodged some loose rock, which fell and gashed you in the right shin. You are bleeding fairly steadily and, while you can walk, you can only do so with significant pain and with about three times as much effort as you normally can.	X	X
Explosion outby. You hear a loud explosion fairly close to your position, outby. You immediately feel an increase in temperature. A call indicates that the team working several crosscuts away from yours reported heavy smoke outby of their position.	X	X
Explosion and broken leg. There was an explosion in the mine coming several crosscuts away. The explosion dislodged some loose rock, which broke a bone in your lower leg—preventing you from putting any pressure on that leg. You notice heavy smoke coming from the direction of the explosion.		X
Explosion and methane. You hear an explosion coming from outby. Soon after, you receive a call from the miners who were working in the area explaining that the explosion damaged the ventilation controls and that methane levels had jumped to over 7% and were climbing when they had evacuated.		X

Note: X's indicate that the corresponding risk was included in the survey to NIOSH and mine inspectors, respectively.

measures compared both scales using only the scenarios that were common in both.

To test for effects in response presentation for the secondary study, respondents were randomly assigned to the different response alternative formats ("3F," "2F," "3C") in a 3-group design. For each respondent, the same format was used across all emergency risk scenarios.

3.1.4. *Procedures*

The survey was administered to all samples online. For the initial pilot NIOSH data, the survey was sent to NIOSH and was disseminated through an e-mail to potential volunteers in November and December, 2010. For the mine inspectors, the revised survey was delivered through NIOSH and administered during a Mine Safety and Health

Table II. Survey 1 Rank Orderings

	Survey 1 NIOSH ($n = 18$)	Survey 1 Inspectors ($n = 56$)
Hurt foot	1	1
Temporary haze	2	–
Break in ventilation	3	3
Burnt-rubber smell	4	4
Roof fall along the face	5	5
Pillar squeezing	–	6
Roof fall outby	6–7	7
Reverse in ventilation		8
Mantrip fire	–	9
Stop in ventilation	8	10
Elevated level of methane	9	11
Smoke outby	10	12
Explosion and hurt shin	11	13
Explosion outby	12	14
Explosion and broken leg	–	15
Explosion and elevated methane	–	16
Kendall's W	0.66	0.43
Average pairwise agreement	82.9%	76.1%

Administration (MSHA) training that occurred from January through March, 2011.

3.2. Results

3.2.1. Kemeny Ranking

The Kemeny ranking for the different pools of respondents are included in Table II. As the rank ordering was optional in the survey, rank orderings were provided by all 18 NIOSH employees, but only 56 of the 61 mine inspectors.

In reviewing the respondent's answers, several mine inspectors appear to have provided rank orderings in the opposite direction of most respondents. For example, six participants (10.5%) were identified as having ranked "hurt foot" as being more risky than "explosion and elevated methane." Our analysis below includes these responses as reported.

3.2.2. Internal Consistency

We measured agreement between respondents using Kendall's W. For NIOSH employees, we found a moderate-high Kendall's W of 0.66 ($p < 0.001$); for the inspectors, we find a moderate W of 0.43 ($p < 0.001$).

We tested for agreement between respondents and the aggregated risk scale using average pairwise agreement. Average pairwise agreement for each scale was generally higher among NIOSH employees at 82.9% compared to the mine inspectors at 76.1%. The pairwise agreement matrix for each sample is provided in Fig. 3, with each cell interpreted as the percentage of respondents ranking the scenario in the row as less risky than the scenario in the column. In general, there was greater consensus in the rank ordering of scenarios with larger aggregate rank differences than those with smaller rank differences. Consensus in the pairwise agreement matrix is indicated by values close to 0% or 100%.

3.2.3. Scale Reliability

We tested the reliability of the scale rank by comparing the two aggregate rank orderings using Kendall's tau-b. The two scales showed a high association, $\tau_B = 0.991$. To test for significance, we used the methodology indicated by Noether,[19] finding that the association was significant, $z_B = 4.22, p < 0.001$.

3.2.4. Scale Validity

We tested the scale's validity by comparing the scale to the reported response to the scenarios. As the option to "Continue Working" appeared across all treatment conditions, we used the proportion of participants choosing to continue work as a measure of how safe (nonrisky) each scenario was. Kendall's tau-b was fairly high for both the NIOSH employees, $\tau_B = 0.83$ ($z_B = 3.68$, $p < 0.001$), and the mine inspectors, $\tau_B = 0.80$ ($z_B = 4.09$, $p < 0.001$). The relationship is diagrammed in Fig. 4, with the percentage of participants choosing to continue working represented on the vertical axis and the scenarios on the horizontal axis, ranked in increasing order according to all responses. As scenarios became more risky toward the right end of the horizontal axis, the proportion of NIOSH employees and miners that decided to continue working went to zero.

3.2.5. Response Presentation Comparisons

We tested for differences in responses as a function of response presentations only for mine inspectors, as the NIOSH sample was very small. The percentage of respondents choosing to continue work was used as the main dependent measure, as it was

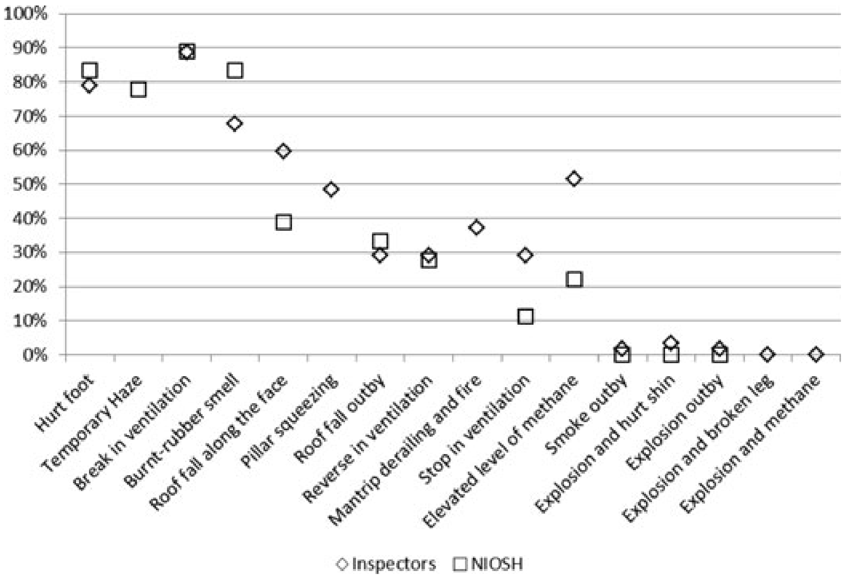

Fig. 4. Percent respondents choosing to continue work by scenario.

shared across all treatments, and was interpreted as the most risky response to take.

We used a nonparametric bootstrap to test for differences in two ways. First, we estimated the percentage of times that the maximum observed difference in the decision to continue working between treatments *across all risk scenarios* was greater than the maximum observed difference in our bootstrap sample in an approach similar to the Kolmogorov–Smirnov test. We created a bootstrap sample by randomly resampling our participant responses, with replacement, from all three treatments into new groups identical in size to our original treatment groups (3F: $n = 23$; C: $n = 20$; 2F: $n = 19$). Then, we identified the maximum difference for a risk scenario across all scenarios in the sample. Repeating this process 10,000 times, we estimated a distribution of maximum differences. The observed maximum difference was in "Stop in ventilation" between the 2F and 3F treatments, with a difference of 29.1%. By examining the percent of times our bootstrap distribution exceeded this observed value, we estimate $p = 59.4$, a nonsignificant difference. A diagram of emergency response is provided in Fig. 5, including reported responses for exiting the mine and seeking a refuge‘ alternative.

Next, we compared the maximum difference in the percentage of respondents who chose to continue working across treatments for each risk scenario against the distribution derived from our bootstrap sample, in an approach similar to running a separate significance test for each scenario. The bootstrap and p-values were calculated, as described earlier, and we find that all scenario comparisons are nonsignificant, except for "Stop in ventilation" with $p = 0.10$. Given no adjustment for multiple comparisons in the second analysis, the significance level for this scenario may be overstated.

3.3. Discussion

3.3.1. Scale Validation Implications

Kendall's W was moderately high for both NIOSH employees and mine inspectors, suggesting fair agreement across individuals and with the aggregated scale. Nonetheless, we should consider reasons for why they were not higher. First, respondents may have interpreted the term "risk" differently—as it was not specified in the survey. For example, risk may be interpreted as including any combination of fatality or injury. Participants might also include nonhealth effects in their interpretation of risk, such as dread or unfamiliarity.[20] Kendall's W being higher for NIOSH employees than for the mine inspectors suggests that safety researchers view risks more consistently, potentially as a result of a more academic and formalized view of risk. Second, the scenarios may have been worded unclearly, leading to unintended and inconsistent interpretations. Finally, the respondents may have genuinely differing opinions on how the risks rank. Unfortunately, given limited access to the mine inspectors, we were

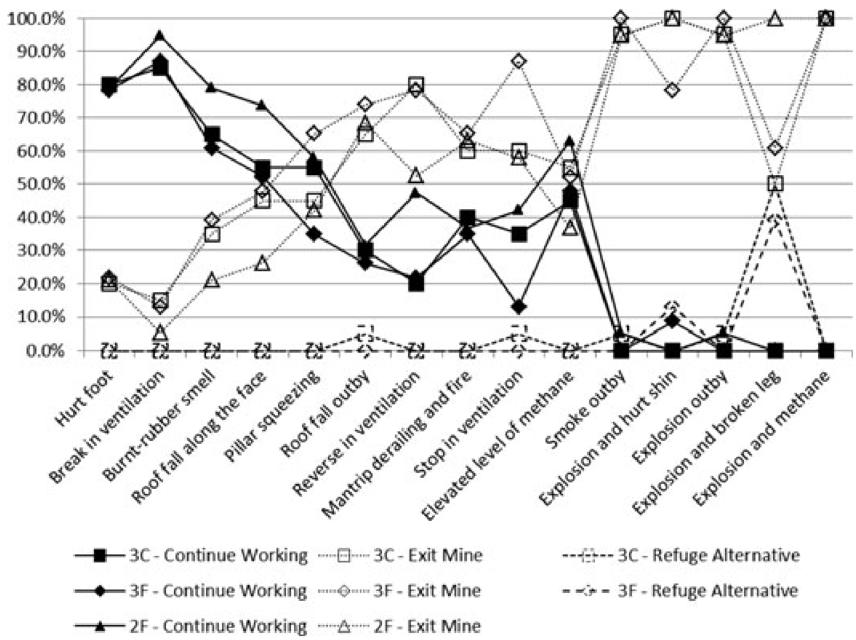

Fig. 5. Emergency response by response presentation and risk scenario.

unable to gather more information about these possible causes.

We did not expect the Kendall's tau-*b* between the two aggregate rank orderings to be as high as it was, given the two different populations responding to the survey. This high association is encouraging, because it suggests that the developed scale can be commonly understood across mine safety officials and miners. The one set of scenarios that are rank ordered differently—a tie for NIOSH employees between "Roof fall outby" and "Stop in ventilation," but not for the mine inspectors—might be explained as a result of the small sample of NIOSH employees or due to true differences in perceived rank orderings. If the latter is true, follow-up interviews between NIOSH and the mine inspectors may determine the reasons for these differences, e.g., differences in knowledge, and to build consensus.

3.3.2. *Scale Applications*

If the risk scale is considered reasonable, the next step would be to apply the scale appropriately for its intended purpose. We suggest three ways this may be done.

In certain cases, the whole scale does not need to be communicated to the target audience. For example, to communicate a precise risk level under which specific emergency responses should be taken, it may be more effective to communicate only those scenarios at that risk threshold. The pairwise agreement matrix can provide information on the effectiveness of such a threshold, giving indications as to the types of risks that are more likely to be confused as falling above or below the threshold and the probability of such confusion. Providing a few similarly risky scenarios could reduce this confusion further.

In other cases, researchers or trainers may wish to communicate distinct, but not precise risk levels, for example, to describe a case under which an emergency response behavior should be evidently appropriate. In these cases, risk scenarios should be selected far from the risk thresholds and the pairwise agreement matrix can be used to determine which risks are viewed as clearly distinct.

Finally, the whole risk scale may be used in cases where general patterns of behavior need to be assessed. This is more likely to be used among researchers, such as in examining whether the effects of a policy affect behavior overall or if researchers are uncertain as to where such effects would occur. Our secondary study on how the impact of emergency response presentation on behavior is an example of such a case.

3.3.3. *Response Presentation*

Given the relatively recent introduction of refuge alternatives to mine emergency response, we wanted to test if the introduction of refuge alternatives and the manner in which they were presented could impact how miners respond to emergency scenarios. Behavior could become riskier, due to moral hazard,[21] if the refuge alternative was seen as a "fall-back" that allowed them to continue under conditions in which they would have otherwise attempted to exit. Miners could also be less risky due to the compromise effect,[22] as exiting the mine now seems like an objectively less extreme response. Similarly, as exiting the mine appears in the center of the response scale, they may interpret it as a more normatively appropriate response.[23]

Our results suggest either that these effects do not have a significant impact or that they interact in such a way as to negate each other's influence. Including refuge alternatives directionally results in less risky behavior, with a presentation on equal footing resulting in the least risky behavior. Given non-significant differences, it is difficult to provide strong recommendations, although, in actual practice, some communication must be made. As refuge alternatives exist, either the 3F or 3C presentation should be used based on which presentation results in the pattern most similar to desired behavior.

4. MINE EMERGENCY RISK SCALE, SURVEY 2

To test some methodological improvements, a second survey was run using a new group of mine inspectors completing their MSHA training. The improvements were focused on reducing the need for follow-up interviews by gathering feedback as part of the survey itself; providing a clearer definition of "risk" for the rank ordering; and implementing checks to ensure that participants had read the rank ordering instructions correctly. As with the initial survey, a secondary study was included—this time to test whether the respondent's perceived role on the crew, as a member of the crew or as the crew foreman, would impact the emergency response alternatives they select.

4.1. Methodology

4.1.1. Participants

Forty-seven mine inspectors were asked to complete the second survey. Respondents worked an average of 30.0 years ($SD = 12.6$) in the mining industry and 25.9 years ($SD = 13.0$) inside a mine. Average age of respondents was 52.9 years ($SD = 10.1$), and none of the respondents were female.

4.1.2. Materials

Most of the survey materials were identical to the first survey, with some targeted modifications. Two versions of the introduction were created, asking the respondent to answer the question either as "a crew member" or as "the crew foreman," and are presented below:

> On the following page, you will be presented with several scenarios that involve some risk of fatality. We ask that you consider these scenarios as if you were [a member/the foreman] of a crew of underground coal miners.
>
> Imagine that your crew just started a shift and are currently at the mine face, approximately 40 crosscuts inby the entrance. A refuge alternative is located 10 crosscuts outby your position. Please take a moment to picture yourself in this situation.
>
> On the next page are scenarios and several actions that you could recommend to your crew. Taking the role of [a crew member/the crew foreman], indicate which action you would recommend. If you would recommend doing something not listed, please select the best answer from the options provided then clarify your response or note what other action you would take in the comments section for that scenario.
>
> Please keep in mind that each scenario should be considered separately.

Respondents were then presented with the same mine risk scenarios in a randomized order and asked how they would respond. Responses were solicited using the 3F format from the first survey, modified with an option to comment on the response, as illustrated in Fig. 6.

Respondents were required to rank order the scenarios using the same drag-and-drop interface; however, the wording for risk was revised to read "most likely to result in a fatality" to "least likely to result in a fatality." After the rank ordering, participants were provided with the scenario that they ranked as most likely and least likely to result in a fatality with the following question: "To check that we recorded your answers accurately, could you confirm that you ranked the scenario <most likely> as more likely to result in a fatality than <least likely>?" The participant then had the option of answering "Yes" and continuing the survey; answering "No, the ranking is reversed" indicating that their rank ordering should be reversed for analysis; or going back to the rank ordering and revising the order.

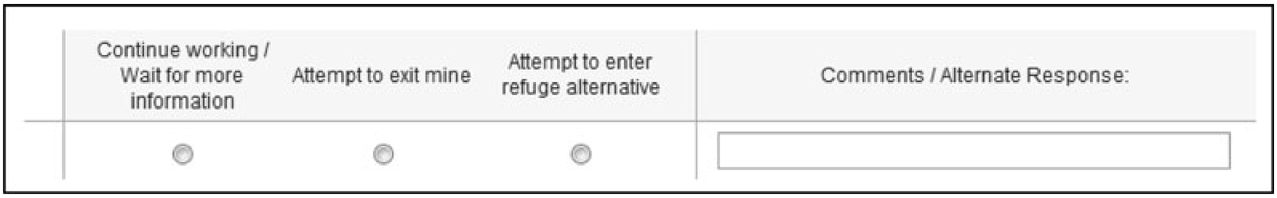

Fig. 6. Response format, Survey 2.

Starting from lowest to highest rank ordered risk, the survey identified the first instance in which the respondent chose to continue working in the less risky scenario but exit the mine in the more risky scenario. The respondent was then prompted to answer the following: "You ranked the scenario <less likely> next to the scenario <more likely>. However, you answered that you would respond to these risks differently. // Why did you decide to continue working for <less likely> but attempt to escape the mine for <more likely>?" The survey presented a similar question for exiting the mine and attempting to enter the refuge alternative. Finally, the survey asked the respondent what factors they would consider before taking each behavioral response, as follows: "In general, what factors do you feel would influence a miner to [continue working/attempt to exit the mine/attempt to enter a refuge chamber]?"

4.1.3. Design

An aggregate risk scale was developed based on the mine inspector responses, with internal consistency and validity measures calculated for the scale. Reliability measures were calculated between the current survey and the first survey.

For the secondary study, respondents were randomly assigned to the different roles ("crew member," "crew foreman") in a two-group design.

4.1.4. Procedures

The survey was administered online to the mine inspectors during MSHA training in July, 2011.

4.2. Results

4.2.1. Kemeny Ranking

The Kemeny ranking of all 47 respondents are included in Table III. Three respondents (6.4%) indicated that they had reverse ranked the scenarios. After correcting those respondents' rank orderings, no respondent were identified to have ranked "hurt foot" as more likely to result in a fatality than "explosion and elevated methane."

4.2.2. Internal Consistency

We found a moderate-high Kendall's W of 0.58 ($p < 0.001$) and an average pairwise agreement of 79.6%. The pairwise agreement matrix for each sample is provided in Fig. 7. In general, there is greater consensus in the rank ordering of scenarios with larger rank differences in risk than those with smaller rank differences.

4.2.3. Scale Reliability

We test the reliability of the scale by comparing the rank orderings to the first survey, using Kendall's tau. The two scales show a high association, $\tau = 0.867$ ($z_B = 4.50, p < 0.001$).

Table III. Survey 2 Rank Orderings

	Survey 2 Inspectors ($n = 47$)
Hurt foot	1
Break in ventilation	2
Burnt-rubber smell	3
Pillar Squeezing	4
Roof fall outby	5
Stop in ventilation	6
Reverse in ventilation	7
Roof fall along the face	8
Mantrip fire	9
Elevated level of methane	10
Smoke outby	11
Explosion and hurt shin	12
Explosion and broken leg	13
Explosion outby	14
Explosion and elevated methane	15
Kendall's W	0.58
Average pairwise agreement	79.6%

	Hurt foot	Break in ventilation	Burnt-rubber smell	Pillar squeezing	Roof fall outby	Stop in ventilation	Reverse in ventilation	Roof fall along the face	Mantrip derailing and fire	Elevated level of methane	Smoke outby	Explosion and hurt shin	Explosion and broken leg	Explosion outby	Explosion and methane
1. Hurt foot		70%	83%	87%	83%	87%	89%	83%	91%	85%	96%	96%	96%	98%	100%
2. Break in ventilation	30%		66%	72%	74%	79%	79%	83%	79%	83%	87%	94%	96%	96%	98%
3. Burnt-rubber smell	17%	34%		60%	60%	49%	49%	68%	74%	70%	96%	91%	94%	98%	98%
4. Pillar squeezing	13%	28%	40%		55%	49%	53%	60%	62%	68%	91%	89%	94%	94%	100%
5. Roof fall outby	17%	26%	40%	45%		53%	60%	55%	57%	62%	94%	91%	94%	96%	98%
6. Stop in ventilation	13%	21%	51%	51%	47%		55%	62%	60%	68%	87%	83%	89%	98%	98%
7. Reverse in ventilation	11%	21%	51%	47%	40%	45%		53%	57%	60%	83%	87%	89%	96%	96%
8. Roof fall along the face	17%	17%	32%	40%	45%	38%	47%		51%	55%	87%	85%	91%	91%	98%
9. Mantrip derailing and fire	9%	21%	26%	38%	43%	40%	43%	49%		49%	89%	83%	87%	94%	98%
10. Elevated level of methane	15%	17%	30%	32%	38%	32%	40%	45%	51%		77%	72%	85%	89%	94%
11. Smoke outby	4%	13%	4%	9%	6%	13%	17%	13%	11%	23%		57%	62%	81%	91%
12. Explosion and hurt shin	4%	6%	9%	11%	9%	17%	13%	15%	17%	28%	43%		81%	70%	91%
13. Explosion and broken leg	4%	4%	6%	6%	6%	11%	11%	9%	13%	15%	38%	19%		55%	85%
14. Explosion outby	2%	4%	2%	6%	4%	2%	4%	9%	6%	11%	19%	30%	45%		79%
15. Explosion and methane	0%	2%	2%	0%	2%	2%	4%	2%	2%	6%	9%	9%	15%	21%	

Fig. 7. Pairwise agreement matrix by scenario, survey.

4.2.4. *Scale Validity*

We test the validity by comparing the scale to the proportion of participants choosing to continue work for each scenario. Kendall's tau-b was moderate-to-high, $\tau_B = 0.75$ ($z_B = 3.83$, $p < 0.001$). The relationship is diagrammed in Fig. 8.

4.2.5. *Differences in Response by Role*

We tested for any effects of role on reported response to the different risk scenarios using the bootstrap methodologies described earlier. Testing the maximum difference in decision to continue working across all scenarios, we find no significant differences ($p = 0.24$). Testing each scenario separately, we find that no difference is significant except for "Roof fall outby," which differs by 35.6% ($p = 0.05$). These results are diagrammed in Fig. 9.

4.3. Discussion

4.3.1. *Rank Order Confirmation*

As the percent of respondents who indicated that they reverse ranked their questions closely matched the estimated number of reverse ranked questions from the first survey, it seems reasonable to assume that at least some of those rankings from the first survey were unintentionally reverse ranked. The rank

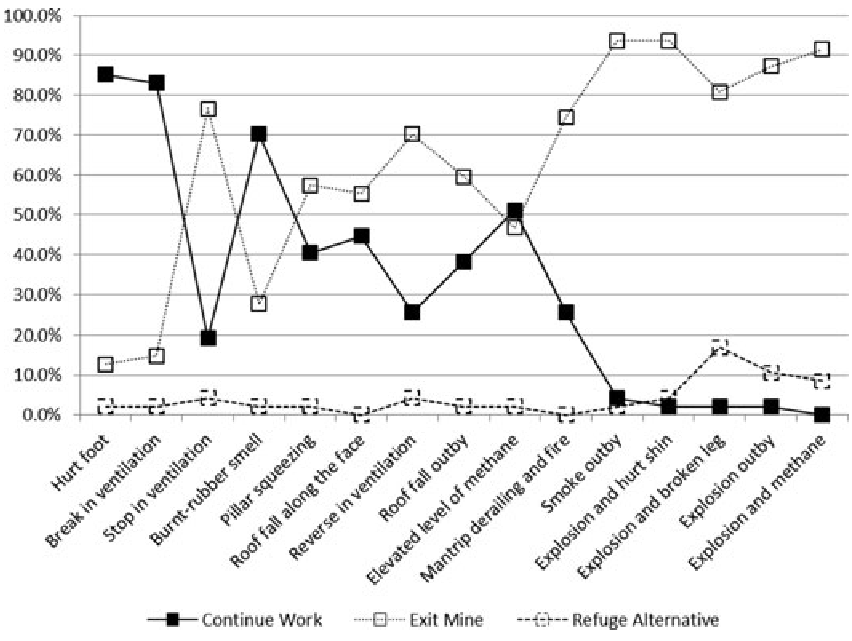

Fig. 8. Emergency response by response presentation and risk scenario.

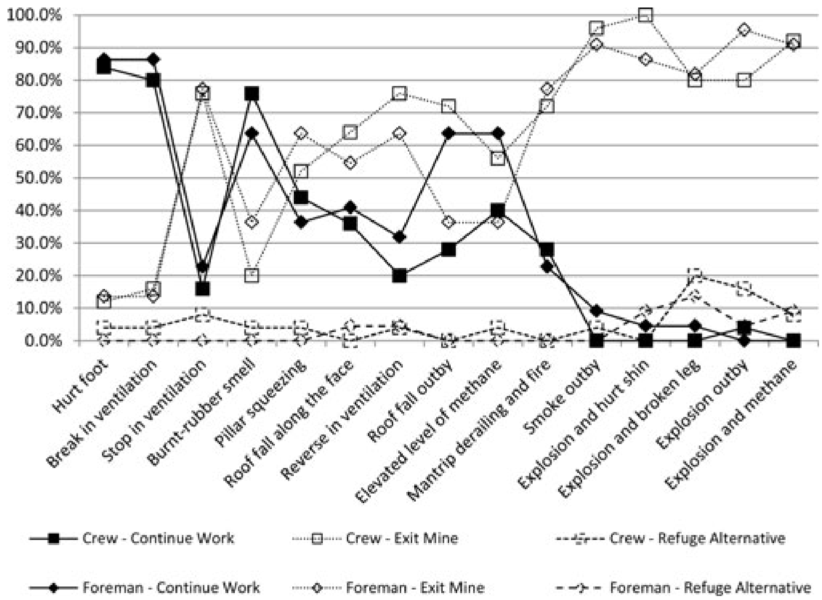

Fig. 9. Emergency response by role, Survey 2.

order confirmation question seems to eliminate the incidence of such errors and is recommended in future surveys.

4.3.2. *Scale Reliability*

Although the Kendall's tau measure of association between the first and second survey is fairly high, there are differences in the rank ordering between the first and second survey. A closer inspection reveals that the more substantial changes, i.e., nonconsecutive rank order shifts, are accounted for by two risks in particular: "Stop in ventilation," which is considered substantially less risky/less likely to result in a fatality; and "Roof fall along the face," which is considered substantially more risky/more likely to result in a fatality. It is possible that the changes in rank ordering reflect underlying noise that challenges the reliability of this methodology; however, the high Kendall's W for both studies suggest this possibility is unlikely. Mine inspector's opinions could have changed in the intervening time period; however, this too seems unlikely given the similar reported emergency responses for both scenarios.

One possible cause of the discrepancy is the change in wording from "most risky"/"least risky" to "most likely to result in a fatality"/"least likely to result in a fatality." In the first survey, respondents may have considered a stop in ventilation risky as it increased the likelihood of other risks (such as increased levels of methane), but in the second survey, it is not considered a direct cause of fatalities. In a similar fashion, a roof fall along the face is less likely than a roof fall outby to threaten the structural integrity of a miner's escape routes; however, is more likely to result in a direct fatality as most miners will be located near the face during their work. The implications of this explanation highlights the potential unintended trade-offs of increasing specificity in defining risk, and it reinforces the need to have a solid understanding of the nature of the risk of interest. The greater association between the risk ranking and emergency response in the first survey compared to the second suggests that a less precise definition of risk can, in some cases, provide a better measure. If due to noise, this reinforces the recommendation to provide a few risk scenarios, rather than a single risk scenario, in communicating risk thresholds.

4.3.3. *Open-Ended Questions*

Open-ended questions were added to survey to allow respondents to indicate any uncertainty they had or to qualify their responses. In general, respondents did not use the open-ended questions to indicate uncertainty, but did provide insights into their perspectives on the risk scenarios.

The comments in the emergency response section predominantly noted actions that the respondents would take to investigate the source of the risk cues or to remediate the risk. That a substantial proportion of participants were concerned about risk remediation suggested another potential source of noise for our scale: whether or not participants implicitly included or excluded the effects of potential remediation in their evaluation of risk.

Toward the end of the survey, there were two sets of questions intended to assess the factors that were used to decide which emergency response to take. The first set of questions focused on consecutively ranked risk scenarios where the respondent chose a more aggressive emergency response in one scenario than in the other. However, as most respondents never selected use of the refuge alternative as an option, only comparisons of continuing work and attempting to exit the mine had sufficient responses for evaluation. Of the 47 responses, 13 (27.7%) mentioned the potential for remediation, nine (19.1%) mentioned restriction of escape ways, six (12.8%) mentioned uncertainty about the cause, four (8.5%) mentioned additional or escalating risks, two (4.3%) mentioned the number of miners affected, and one (2.1%) mentioned regulatory requirements.

The second set of questions asked respondents for factors that generally contribute to them taking a specific emergency response. Most of the response regarding continuing work or attempting to escape noted risk scenarios, e.g., fires or explosions, rather than more basic factors—however, the responses regarding refuge alternative use were somewhat more sophisticated. Of the 45 responses, 40 (88.9%) mentioned the ability/inability to escape, six (13.3%) mentioned excessively high or increasing danger, four (8.9%) mentioned physical injury, three (6.7%) mentioned that refuge alternatives should never be used, two (4.4%) mentioned accessibility/prior training in the use of the refuge alternatives, and two (4.4%) mentioned individual personalities.

As noted at the beginning, one of the first tasks in developing a scenario-based risk scale is assessing the nature and scope of the risks being addressed. Inclusion of such open-ended questions provides a way to assess how respondents are approaching the question of risk more generally. The complexity in which risk is perceived by miners, revealed by these open-ended questions, provides additional evidence that a broader ranking by "risk" rather than "risk of fatality" may more comprehensively capture the nature of the risks being examined.

4.3.4. Role Differences

Our results do not indicate a significant difference in emergency response behavior depending upon prescribed role, with the possible exception of "Roof fall outby." Nonetheless, in 10 of 15 scenarios, the foreman role appears directionally more risk-seeking than the crew member role.

5. GENERAL DISCUSSION

Effective emergency response training requires the communication of commonly understood risk levels, which can be applied in dynamic scenarios with as little noise as possible. In this article, we proposed a methodology to develop a context-specific, scenario-based risk scale comprised of a Kemeny ranking of risk scenarios developed in conjunction with the target training population, as well as several suggested measures to assess the scale's internal consistency, reliability, and validity. The second survey focused on introducing practical improvements to reduce response errors and to capture additional feedback from the target audience. We present this methodology as new, suggesting that there is opportunity for broader applications and improvement. Moreover, developing a scenario-based risk scale is only a part of the effort required in emergency response training. In closing, we provide a brief review of where the scenario-based risk scale falls in line with other work, and suggest further lines of improvements and research.

5.1. Part of a Larger Process

As we noted in the Introduction, emergency response scenarios often differ from other decision-making environments due to a dynamically changing environment and to novel risk cues. In cases such as mine emergencies, decisions often need to be made in teams and quick agreement among all decision makers may be important to respond effectively. To that effect, we have focused on *how to communicate* appropriate risk thresholds, while keeping relatively silent on *how to establish* these risk thresholds.

We note several potentially complimentary approaches in how to establish risk thresholds. If a quantitative assessment of risks were available, we may be able to use contingent valuation methodologies to establish risk thresholds. If these risks could then be aligned with the scenario-based risk scales, e.g., through simulations, historical data, or expert estimates, we could translate the quantitative risk thresholds to scenarios for training and communication.

Discussing emergency risks in a structured but more qualitative manner with the target population may also help us to better identify what matters to them and how much it matters.[24] Our efforts in providing and analyzing the open-ended questions as part of the second survey acts as an extension of such efforts, allowing us to better understand the factors that influence mine emergency response which often are not restricted purely to loss of life or risk. However, the open-ended questions in the survey compliment but do not substitute for a more rigorous program of understanding the target population's preferences as can be obtained through more flexible procedures, such as a guided interview.

5.2. Further Research

While we proposed the scenario-based risk scale as an alternative to quantitative scales or "familiar" risk scales, our article does not provide a direct contrast of the methodologies. This is partly to the credit of the scenario-based risk scale, as it allows for the development of a risk scale in cases like mine emergency response, where concrete quantitative risk estimates do not exist. Nonetheless, it would be valuable to test these methodologies more directly in domains where quantitative estimates are available.

Additionally, while we performed a basic analysis of the difference in rank orderings in the first survey to determine that some of the rankings appeared reversed, it may be valuable to develop methodologies to evaluate whether or not there is any clustering of risk rankings. A cluster analysis approach can provide further information regarding how respondents perceive the risk and provide more focused direction on how to reduce ambiguity in the scenarios' description.

Finally, although we remarked on the multidimensionality of emergency response; our approach so far has been to "encompass" those dimensions under a single risk construct. A contrasting approach would be to solicit several rank orderings of the scenarios based on more specific dimensions of the construct, e.g., asking one sample to rank order by "likelihood of fatality," another by "likelihood of injury," another by "likelihood of blocked escape," etc. Such specification may improve the internal consistency of the scales—as was observed between surveys 1 and 2. When it is inadvisable to communicate the full rank orderings by all dimensions to the target population, methods to collapse the scales after they are established may be identified or communications can simply be designed to refer only to relevant scenarios and dimensions.

ACKNOWLEDGMENTS

Work completed while all authors were at the Dynamic Decision Making Laboratory at Carnegie Mellon University. The authors are thankful to Stephen Broomell and Hau-yu Wong for comments. This research was partially supported by the National Institute for Occupational Safety and Health (Centers for Disease Control and Prevention, Award number: 200–2009-31403) award to Professor Cleotilde Gonzalez.

REFERENCES

1. Hall EA, Margolis KA. Emergency Escape & Refuge Alternatives Instructor Guide and Lesson Plan. Pittsburgh, PA/Spokane, WA: Department of Health and Human Services (National Institute for Occupational Safety and Health), Report No.: 2011-101, October, 2010.
2. Vaught C, Hall EA, Klein KA. Harry's Hard Choices: Mine Refuge Chamber Training Instructor's Guide. Pittsburgh, PA: Department of Health and Human Services (National Institute for Occupational Safety and Health), Information Circular 9511, March 2009.
3. Ball LK, Evans G, Bostrom A. Risky business: challenges in vaccine risk communication. Pediatrics, 1998; 101(3):453–458.
4. Paling J. Strategies to help patients understand risks. BMJ, 2003; 327(7417):745–748.
5. Slovic P, Flynn JH, Layman M. Perceived risk, trust, and the politics of nuclear waste. Science, 1991; 254(5038):1603–1607.
6. MacGregor DG, Slovic P, Morgan MG. Perception of risks from electromagnetic fields: A psychometric evaluation of a risk-communication approach. Risk Analysis, 1994; 14(5):815–828.
7. Wright WF, Urton A. Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. Organizational Behavior and Human Decision Processes, 1989; 44(1):68–82.
8. Lerner JS, Gonzalez RM, Small DA, Fischhoff B. Effects of fear and anger on perceived risks of terrorism. Psychological Science, 2003; 14(2):144–150.
9. Parker AM, Fischhoff B. Decision-making competence: External validation through an individual-differences approach. Journal of Behavioral Decision Making, 2005; 18(1): 1–27.
10. Fielder K. Causal schemata: Review and criticism of research on a popular construct. Journal of Personality and Social Psychology, 1982; 42(6):1001–1013.
11. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: Frequency formats. Psychological review, 1995; 102(4):684–704.
12. Cuite CL, Weinstein ND, Emmons K, Colditz G. A test of numeric formats for communicating risk probabilities. Medical Decision Making, 2008; 28(3):377.
13. Schapira MM, Nattinger AB, McHorney CA. Frequency or probability? A qualitative study of risk communication formats used in health care. Medical Decision Making, 2001; 21(6):459–467.

14. Woloshin S, Schwartz LM, Byram S, Fischhoff B, Welch HG. A new scale for assessing perceptions of chance: A validation study. Medical Decision Making, 2000 ; 20(3):298–307.
15. Connelly NA, Knuth BA. Evaluating risk communication: Examining target audience perceptions about four presentation formats for fish consumption health advisory information. Risk Analysis, 1998; 18(5):649–659.
16. Slovic P, Kraus N, Covello VT. What should we know about making risk comparisons? Risk Analysis, 1990; 10(3):389–392.
17. Young HP. Condorcet's theory of voting. The American Political Science Review, 1988; 82(4):1231–1244.
18. Levin J, Nalebuff B. An introduction to vote-counting schemes. The Journal of Economic Perspectives, 1995; 9(1):3–26.
19. Noether GE. Elements of Nonparametric Statistics. New York: John Wiley & Sons, 1967.
20. Slovic P. Perception of risk. Science, 1987; 236(4799):280–285.
21. Pauly MV. The economics of moral hazard: Comment. The American Economic Review, 1968; 58(3):531–537.
22. Simonson I. Choice based on reasons: The case of attraction and compromise effects. Journal of Consumer Research, 1989; 16(2):158–174.
23. Schwarz N, Hippler H-J, Deutsch B, Strack F. Response scales: Effects of category range on reported behavior and comparative judgments. Public Opinion Quarterly, 1985; 49(3):388–395.
24. Fischhoff B. Giving advice. Decision theory perspectives on sexual assault. The American psychologist, 1992; 47(4):577–588.